INVESTIGATING THE ROLE OF A NOVEL EPIGENETIC MODIFIER, REARRANGED L-MYC FUSION, IN THE MOUSE

Lauren M Bourke

06558607

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Epigenetics Laboratory,

QIMR Berghofer Medical Research Institute

Institute of Health and Biomedical Sciences,

Translational Research Institute,

School of Biomedical Sciences, Faculty of Health

Queensland University of Technology

2016

Abstract

Background

Rearranged L-Myc Fusion, *Rlf,* was recently identified as a novel epigenetic modifier from a mouse N-ethyl-N-nitrosourea mutagenesis screen. The mice used in this study carry a multi-copy green fluorescent protein (GFP) transgene linked to an erythroid specific α-globin promoter that is sensitive to epigenetic silencing. Three independent mouse lines, *MommeD8, MommeD28* and *MommeD34,* with mutations in *Rlf* were each found to have decreases in the percentage of red blood cells expressing GFP, suggesting *Rlf* acts an epigenetic modifier and plays a role in transcriptional activation. Our study is the first to reveal a role for *Rlf* in epigenetics, and currently very little is known about the function of Rlf. These are the first available mouse mutants that can be utilised to study the molecular and phenotypic consequences of *Rlf* inactivation.

Results

Using a combination of RNA sequencing, chromatin immunoprecipitation sequencing and whole genome bisulphite sequencing, I have shown Rlf acts as an epigenetic modifier. Rlf plays a critical role in heart development as determined by histology and this may be mediated via alteration to the Notch signalling pathway, as determined by RNA sequencing. These effects do not appear to be accumulated trasngenerationally.

Conclusion

The results presented in this thesis indicate *Rlf* plays an important role during development and inactivation of *Rlf* alters the transcriptome, methylome and phenotype in the mouse.

Keywords

Epigenetics, gene expression, heart development, methylation, Rearranged L-Myc Fusion, transcription, transgenerational epigenetics.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Ab1-3	Rlf custom antibody 1-3
ac	Acetylation
ADP	Adenosine diphosphate
AGRF	Australian genome research facility
Ald1a7	Aldehyde dehydrogenase family 1, subfamily a7
Apoc1	Apolipoprotein C-I
ar1	Mono- ADP ribosylation
arn	Poly - ADP ribosylation
AS	Angelman syndrome
Atp2b2	ATPase, Ca++ transporting, plasma membrane
ATPase	Adenosine triphosphatase
A^{vy}	Agouti viable yellow
Basp1	Brain abundant, membrane attached signal protein 1
BCA	Bicinchoninic acid
BLAST	Basic local alignment search tool
bp	Base pairs
BSA	Bovine albumin serum
BWA	Burrows-wheeler alignment
CD	Chromodomain
cDNA	Complementary DNA
Cenp-A	Centromere protein A
CHD	Chromodomain helicase DNA-binding
ChIP	Chromatin immunoprecipitation
ChIPseq	Chromatin immunoprecipitation sequencing
CoA	Coenzyme A
Co-IP	Co-immunoprecipitation
co-REST	Rest co-repressor 1
CpG	Cytosine phosphate guanine

Crispld2	Cysteine rich secretory protein LCCL domain containing 2
CSD	Chromoshadow domain
DEPC	Diethylpyrocarbonate
DHS	DNase Hypersensitive Site
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
Dnmt1-3	DNA methyltransferase 1-3
dNTP	Deoxyribonucleotide triphosphate
Dot1l	Dot1-like histone H3K79 methyltransferase
dsRNA	Double stranded RNA
ECG	Echocardiography
EDTA	Ethylenediaminetetra-acetic acid
Ehmt2	Euchromatic histone lysine n-methyltransferase 2
ENCODE	Encyclopedia of DNA elements
ENU	N-ethyl-n-nitrosourea
FACS	Fluorescence activated cell sorting
Fkbp1a	FK506 binding protein 1a
GFP	Green fluorescent protein
GWBS	Genome-wide bisulphite sequencing
H	Histone
H&E	Haematoxylin and eosin
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
Hp1α	Heterochromatin protein 1a
Hpd	4-hydroxyphenylpyruvic dioxygenase
IAP	Intracisternal A particle
Igf2	Insulin-like growth factor 2
Igf2r	Insulin-like growth factor 2 receptor
IHC	Immunohistochemistry
INO80	Inositol requiring 80
IPTG	Isopropyl-thio-β-galactosidase
ISH	In-situ hybridisation

ISWI	Imitation switch
IVS	Interventricular septum
JZ	Junctional zone
K	Lysine
kDa	Kilo Dalton
Kdm6a	Lysine (k)-specific demethylase 6a
Kmt2d	Lysine (k)-specific methyltransferase 2d
lncRNA	Long non coding RNA
Lsd1	Lysine specific demethylase 1
LTR	Long terminal repeat
LV	Left ventricle
LVID	Left ventricle internal diameter
LVPW	Left ventricle posterior wall
LZ	Labyrinth zone
mb	Mega base
MD	Maternal decidua
me	Methylation
Mecp2	Methyl CpG Binding Protein 2
Mgam	Maltase-glucoamylase
min	Minutes
miRNA	Micro RNA
MommeD	Modifier of murine metastable epiallele
mRNA	Messenger RNA
Mycl1	V-myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog
Myh6	Myosin, heavy polypeptide 6, cardiac muscle, alpha
Myh7	Myosin, heavy chain 7, cardiac muscle, beta
Myo5c	Myosin vc
NAD	Nicotinamide adenine dinucleotide
Nkx2-5	NK2 homeobox 5
ns	Not significant
nt	Nucleotide
NuRD	Nucleosome remodelling and deacetylase
PARP	Poly-(ADP-ribose) polymerase

PAS	Periodic acid schiff
PBAF	Polybromo associated factor
Pbrm1	Polybromo 1
PBS	Phosphate buffered saline
PBT	Phosphate buffered saline with tween 20
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
ph	Phosphorylation
PHD	Plant homeodomain
piRNA	Piwi interacting RNA
PIWI	P-element induced wimpy testis
PMSF	Phenylmethylsulfonyl fluoride
Pogz	Pogo transposable element with Znf domain
Ppar-γ	Peroxisome proliferator-activated receptor gamma
Prss50	Protease, serine 50
Psma8	Proteasome subunit alpha type 8
PWS	Prader-Willi syndrome
qRT-PCR	Quantitative real-time reverse-transcription polymerase chain reaction
RIN	RNA integrity number
RISC	RNA induced silencing complex
Rlf	Rearranged L-myc fusion
Rlf-DMR	Rlf differentially methylated region
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
rpm	Revolutions per minute
rt	Room temperature
s	Seconds
SANT	Swi3, Ada2, N-Cor And TFIIIb Domain
SCLC	Small cell lung cancer
SDS	Sodium dodecyl sulfate (lauryl sulfate sodium salt)

SEM	Standard error of the mean
SET	Su(Var)3-9 and enhancer of zeste
siRNA	Short interfering RNA
Sirt6	Sirtuin 6
Smad3/6	Smad family member 3/6
Smarca	Swi/Snf Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A
Smchd1	Structural maintenance of chromosomes flexible hinge domain containing 1
SOB	Super optimal broth
su	Sumoylation
SWI/SNF	Switch/sucrose non-fermentable
TAE	Tris, Acetic Acid and EDTA Buffer
TBE	Tris, Boric Acid and EDTA Buffer
Tbx5	T-box 5
TET	Ten eleven translocation enzyme
TFB I	Transformation buffer I
TFB II	Transformation buffer II
Timd2	T-cell immunoglobulin and mucin domain containing 2
tsDMR	Tissue specific differentially methylated region
TSS	Transcription start site
U	Units
ub	Ubiquitylation
Ube3a	Ubiquitin protein ligase e3a
Uhrf1	Ubituitin like with phd and ring finger domains
V	Volts
Vldlr	Very low density lipoprotein receptor
X-gal	5-bromo-4-chloro-3-indoyl-d-galactoside
Xic	X chromosome inactivation centre
Xist	X inactive specific transcript
Zfp462	Zinc finger protein 462
βactin	Beta-actin gene

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature:

Date: 8th June 2016

Acknowledgements

Even though the size of our lab has reduced each year during my PhD there are still many people I'd like to thank. Firstly, I'd like to thank my supervisor Sarah for being a great mentor and boss, and for taking me on as a student. I have learnt an extraordinary amount over the last 4 years thanks in part to your open door and happiness to talk about anything and everything when I turned up at it. Research's loss is definitely medicines gain, but maybe you'll come back to the 'light side' one day. I'd also like to thank my other supervisors Pam, KumKum and Adrian, who have brought immensely valuable 'non-epigenetics' viewpoints to my project which has incredibly helped my thought process for the better. To Glen and Graham who sat through all of QIMRB reviews and provided sound advice over the three years. Particularly Graham who opened his office door to me whilst Sarah was on maternity leave, and let me borrow any of his developmental text books whenever I needed them.

A huge thanks to the past Whitelaw lab and Harten lab members, especially Vandha and Tam who helped keep the lab such a happy environment and were happy to do the occasional plug check or genotyping PCR for me. To Harry for providing sound bioinformatic advice and not being annoyed by my naïve emails; and Emma because without her *Momme* screen none of this project would have been possible.

Also to all of the experts in their fields that have responded to my emails when I thought I saw something developmentally wrong but didn't really know what I was

looking at, Richard Harvey, Gonzalo Del Monte Nieto, David Simmons, and Dagmar Wilhelm, who were all kind enough to spend their own time providing advice and pointed me towards key research papers. Also to all of the QIMRB staff who helped out with this project: Paul and Sally in sequencing who never got annoyed at my emails asking where my sequencing results were; Clay and Glynn in histology who sat through countless talks about how to best orientate embryos for sectioning; and Ben in the animal house for some great conversations during long weaning sessions.

Thanks to all of the other PhD students at QIMRB, who reminded me that you also need a life outside of the lab, beer and pizza on Fridays, and that indoor netball is an incredibly violent sport and good way to let off steam (maybe next season the Takedowns will win the big cups!).

Lastly a big thanks to all of my family. Aunty Lyn and Uncle John for making sure I ate at least one good meal a week and that I got out of the house and the lab, and providing some incredible pep talks before big events. To my sisters for understanding that I'll always be the favourite (but I did have a twelve year head start), and keeping me updated on all things Darwin and One Direction. But mostly to Mum and Dad who were always happy to take my phone calls on my walks home, and talk me through broken car windscreens, flat tyres and blocked washing machines. All of your support has been greatly appreciated and if I ever win the lotto I'll pay you back.

Publications Arising From This Thesis

Harten, S. K.**,** Oey, H., **Bourke, L. M.**, Bharti, V., Isbel, L., Daxinger, L., Faou, P., Robertson, N., Matthews, J. M., Whitelaw, E**.** (2015). The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers. *BMC Biology, 13*(1), 21. doi: 10.1186/s12915-015-0128-2

Chapter 1: Introduction

1.1 EPIGENETIC MODIFICATIONS

Epigenetics refers to the study of changes in gene expression that are heritable and occur without changes in the underlying DNA sequence, as reviewed in (Dupont et al., 2009). Epigenetic modifiers are genes that are involved in establishing and maintaining the epigenetic state of the genome. They are known to play critical roles in development and disease (Chai et al., 2005; Chowdhury et al., 2011; Howard et al., 2007; Li et al., 1992; Stopka & Skoultchi, 2003). Epigenetic mechanisms that regulate gene expression include DNA methylation, histone and nucleosome modifications, chromatin remodelling protein complexes and non-coding RNA (Allis et al., 2007). The basic repeating unit of chromatin is the nucleosome, which consists of 147 base pairs of DNA wrapped around an octameric histone core containing two copies of histones H2A, H2B, H3 and H4 (Kornberg & Thomas, 1974; Luger et al., 1997). Chromatin is then packaged either into euchromatin, loosely packed chromatin associated with gene activation, or heterochromatin tightly packed chromatin associated with gene silencing, Figure 1.1 (Jones et al., 2008). Although the importance of epigenetic modifications is now clear, the mechanisms by which these modifications are established and maintained is still not fully understood.

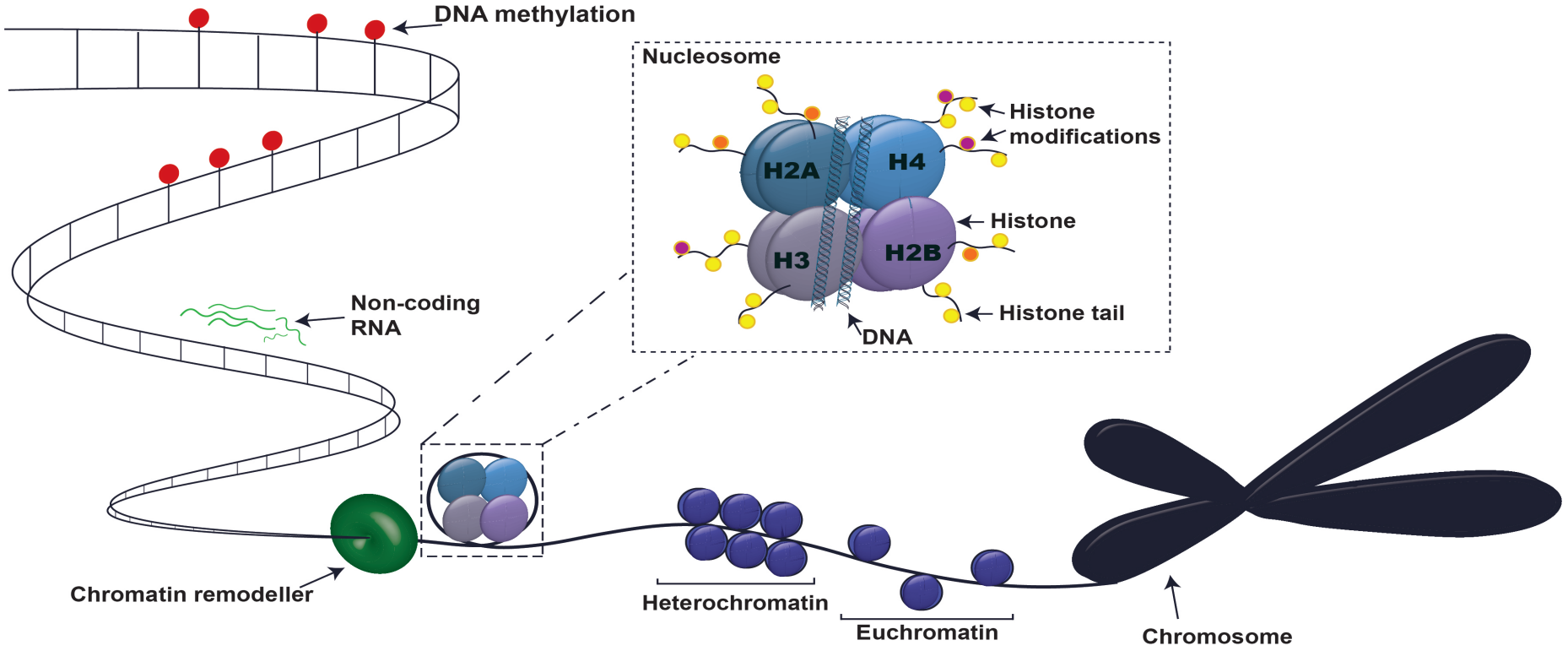


Figure 1.1: A schematic diagram of epigenetic modifications.

Modifications such as DNA methylation, histone modifications, chromatin remodelers and non-coding RNA can alter gene expression. The nucleosome (inset) is the basic repeating unit of chromatin, consisting of DNA wrapped around an octameric histone core containing two copies of the histones H2A, H2B, H3 and H4; chromatin is then packaged into either heterochromatin or euchromatin.

1.1.1 DNA methylation

DNA methylation primarily occurs at the 5' position of a cytosine located next to a guanine (CpG dinucleotides), but may also occur in a non-CpG context as has been reported in mouse and human embryonic stem cells (Lister et al., 2009; Ramsahoye et al., 2000). This review will focus upon CpG methylation, which has been studied the most extensively. The presence of DNA methylation is commonly associated with gene silencing. There is increasing evidence of silencing occurring prior to methylation (Lock et al., 1987; Ohm et al., 2007; Schlesinger et al., 2007).

DNA methylation

There are three DNA methyltransferases (DNMTs) responsible for assembling methylation marks, the maintenance methyltransferase DNMT1 and the *de novo* methyltransferases DNMT3A and DNMT3B (Lei et al., 1996; Okano et al., 1999; Okano et al., 1998). DNMT1 is recruited to hemimethylated DNA by UHRF1, ubiquitin-like with PHD and ring finger domains 1, and then methylates the CpG site on the opposite strand. This process allows for DNA methylation patterns to be maintained following DNA replication (Bestor, 1992; Bostick et al., 2007). In comparison DNMT3A and DNMT3B act as *de novo* methyltransferases, establishing new DNA methylation patterns that are important in development (Okano et al., 1999).

DNA demethylation

DNA demethylation has been observed following fertilisation of the zygote and during the development of primordial germ cells (Mayer et al., 2000; Oswald et al., 2000). DNA demethylation may be passive, whereby newly synthesised DNA is not methylated following replication due to the absence of functional DNMT1 (Oswald et al., 2000). Alternatively, DNA demethylation may occur through the active erasure of methylation marks, the mechanism of which was unknown until recently following the identification of TET (Ten-eleven translocation) family of enzymes (Ito et al., 2011; Tahiliani et al., 2009). TET1-3 proteins have the ability to convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which may then be reverted to unmodified cytosine through replication dependent depletion of 5hmC (Ito et al., 2011; Tahiliani et al., 2009; Valinluck & Sowers, 2007). Alternatively, 5hmC can be further converted to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) which can be recognised by thymine DNA-glycosylase and undergo base excision repair through the DNA repair pathway, resulting in the regeneration of unmodified cytosine (He et al., 2011; Nabel et al., 2012). More recently it has been shown active demethylation by TET3 in the zygote, does not require thymine DNA-glycosylase activity, but an alternate mechanism downstream is occurring (Guo et al., 2014). Although whether this downstream mechanism is the decarboxylation of 5caC, dehydroxymethylation of 5hmC or base excision repair is still unknown (Chen et al., 2012; Hajkova et al., 2010; Schiesser et al., 2012).

1.1.2 Histone modifications

Histones can undergo post translational modifications that lead to alterations in chromatin conformation, and in turn, gene expression, as reviewed in (Strahl & Allis, 2000). This review will focus upon methylation, acetylation and phosphorylation, however other histone modifications also include ubiquitylation, sumoylation and adenosine diphosphate ribosylation. **Table 1.1** lists the nomenclature for describing histone amino acid modifications, as reviewed in (Xhemalce et al., 2006).

Table 1.1 Histone amino acid modification nomenclatures.

Modification	Amino acid modified	Single letter amino acid	Degree of modification	Abbreviation	Example of modification
Acetylation	Lysine	K	Mono	ac	H3K56ac
Methylation	Lysine	K	Mono	me1	H3K4me1
	Lysine	K	Di	me2	H3K4me2
	Lysine	K	Tri	me3	H3K4me3
	Arginine	R	Mono	me1	H3R2me1
	Arginine	R	Di-symmetric	me2s	H3R2me2s
	Arginine	R	Di-asymmetric	me2a	H3R2me2a
Phosphorylation	Serine or Threonine or Tyrosine	S/T/Y	Mono	ph	H3S10ph
Ubiquitylation	Lysine	K	Mono	ub1	H2AK119ub1
Sumoylation	Lysine	K	Mono	su	H2BK6su
ADP ribosylation	Glutamate	E	Mono	ar1	H2BE2ar1
	Glutamate	E	Poly	arn	H2BE2arn

Table 1.1: Histone modifications are written beginning with the name of the histone (e.g. H3 for histone 3), followed by the single letter amino acid abbreviation and its position within the protein (e.g. K56 for lysine 56), and the abbreviated modification (e.g. ac for acetylation). Adapted from Xhemalce et al., 2006, review.

Histone methylation

Histone methylation has been shown to occur on arginine, lysine and histidine residues. Lysine methylation has been well studied and this residue can undergo mono- (me1), di- (me2) or tri- (me3) methylation (Rosenfeld et al., 2009). Whilst methylation of histone 3 lysine 4 (H3K4), H3K36 and H3K79 are associated with transcriptional activation, methylation of H3K9 and H3K27 are associated with transcriptional repression (Rosenfeld et al., 2009). The presence of H3K9me3 in particular has been identified as a hallmark of heterochromatin and thus gene silencing (Peters et al., 2003). Lysine residues are able to be methylated by SET-domain containing and DOT1L-like (DOT1-like histone H3K79 methyltransferase) proteins, whilst they can be demethylated by amine oxidases and jumonji C domain containing proteins (Feng et al., 2002; Rea et al., 2000; Shi et al., 2004; Tsukada et al., 2006). Although, in most cases, it is not yet known whether the addition or removal of chromatin marks are the cause or consequence of changes to chromatin structure.

Although histone methylation is carried out by different proteins to those participating in DNA methylation, the presence of one modification may influence the other (Tjeertes et al., 2009). For example, studies have shown the histone methyltransferase G9a catalyses methylation of H3K9, recruits DNMT3A and DNMT3B, enabling *de novo* DNA methylation and resulting in gene inactivation (Feldman et al., 2006). Separate studies have also shown DNA methylation directs the methylation of H3K9 and prevents H3K4 methylation, leading to a closed chromatin state (Eden et al., 1998; Hashimshony et al., 2003). Other research has shown that knockdown of specific SET (suppressor of variegation, enhancer of zeste

and trithorax) domain histone methyltransferases in plants and fungi results in a loss of DNA methylation at specific regions (Jackson et al., 2002; Tamaru & Selker, 2001).

Histone acetylation

Acetylation of lysine residues is regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs). The presence of acetylation is commonly associated with active transcription, whilst the absence of this modification is associated with transcriptional repression (Kouzarides, 2007). HAT enzymes are able to modify multiple lysine residues, utilising acetyl CoA to transfer an acetyl group to the lysine. This neutralises the positive charge, weakening the interaction between the histone and DNA, enabling transcription, as reviewed in (Yang & Seto, 2007). Whilst most histone modifications occur on the histone tails, acetylation has also been observed at a lysine residue within the histone core. The lysine H3K56 points towards the DNA major groove within the nucleosome and acetylation is believed to alter histone/DNA interactions similar to acetylation of N-tail lysines (Tjeertes et al., 2009).

There are four classes of HDACs based upon their structure, sequence and domain organisation. Class I comprises HDACs 1, 2, 3 and 8. These proteins are ubiquitously expressed and predominantly localised to the nucleus (Taunton et al., 1996). Class II includes HDACs 4, 5, 6, 7, 9 and 10. These HDACs have restricted expression patterns and have the ability to move between the cytoplasm and the nucleus (McKinsey et al., 2000; Vega et al., 2004). Class III HDACs, also known as sirtuins, rely upon NAD+ (nicotinamide adenine dinucleotide) for their activity.

Sirtuins are ubiquitously expressed and can deacetylate histone and non histone substrates (Frye, 1999). Class IV, containing only HDAC11 is expressed in the brain, heart, muscle, kidney and testis, but little is currently known about its function (Gao et al., 2002; Liu et al., 2008).

Histone phosphorylation

While a large number of studies have analysed histone acetylation there is still much to be learned about histone phosphorylation. Histone phosphorylation can occur on serines, threonines and tyrosines, primarily located on histone tails, and is commonly found alongside other histone modifications (Fischle et al., 2005; Xhemalce et al., 2006). For example, the presence of H3S10ph next to H3K9me3 disrupts HP1α/chromatin binding during M phase of the cell cycle (Fischle et al., 2005). Histone phosphorylation is believed to have similar effects on transcription as histone acetylation, one example being the presence of phosphorylation on H3T118 reducing DNA/histone binding, and increasing DNA accessibility to transcriptional machinery (North et al., 2011).

1.1.3 Histone variants

There are variants for each of the core histones as well as the linker histone H1, which binds to the nucleosome at the DNA entry-exit point (Vignali & Workman, 1998). Replacement of the histone proteins with histone variants has been shown to alter nucleosome stability, chromatin formation and also to play roles in DNA repair (Chakravarthy & Luger, 2006; Yoda et al., 2000). Histones H2A and H3 have diversified the most in eukaryotic evolution and will be discussed here (Malik & Henikoff, 2003; Thatcher & Gorovsky, 1994).

There are five histone H2A variants in humans; these include H2A.Z, H2A.X, macroH2A.1, macroH2A.2 and H2A.Bbd. H2A.Z is conserved throughout evolution and has multiple functions, it is associated with active transcription and sites undergoing histone exchange, and localises to transcriptional start sites (TSS) (Barski et al., 2007; Hatch & Bonner, 1988; Jin et al., 2009; Zlatanova & Thakar, 2008). H2A.Z containing nucleosomes are also less stable than nucleosomes containing canonical H2A due to structural differences altering the H3/H4 docking domain (Suto et al., 2000). H2A.X has been found to play an important role in genome stability. Following DNA damage, H2A.X is phosphorylated by DNA dependent protein kinases. This phosphorylation is required to recruit DNA damage repair machinery (Paull et al., 2000; Rogakou et al., 1999). Unlike H2A.Z the presence of macroH2A variants give rise to more stable nucleosomes and are associated with transcriptional repression (Chakravarthy & Luger, 2006; Changolkar & Pehrson, 2002; Muthurajan et al., 2011). H2A.Bbd has only been found in mammals to date, and displays similar properties to H2A.Z. This variant localises to TSS, is associated with transcriptional activation, and inclusion of H2A.Bbd into the nucleosome creates a less stable nucleosome due to disruption of the H3/H4 docking domain (Gautier et al., 2004; Soboleva et al., 2012; Tolstorukov et al., 2012).

The histone H3 variants - H3.1, H3.3 and CENP-A (centromere protein A) each have differing characteristics. H3.1 is commonly associated with repressive histone marks, whilst H3.3 is associated with active transcription (Hake & Allis, 2006; Mito et al., 2005). H3.3 has also been found to replace H3.1 at active genes and co-localise with H2A.Z at the promoters of active genes (Jin et al., 2009). In

comparison to these two variants CENP-A replaces the canonical H3 at the centromere and is required for maintenance of the centromere (Yoda et al., 2000).

1.1.4 Chromatin remodelling complexes

Chromatin remodelling complexes utilise ATP hydrolysis to restructure nucleosomes via sliding, ejection or repositioning (Dai et al., 2009; Kitagawa et al., 2011; Stopka & Skoultchi, 2003). The four families of chromatin remodelling complexes are SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch), CHD (chromodomain, helicase, DNA binding), and INO80 (inositol requiring 80).

SWI/SNF complexes are large, containing 8-14 subunits and a catalytic subunit, SMARCA4 or SMARCA2 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily A, members 4 and 2) (Mohrmann & Verrijzer, 2005). ISWI complexes are smaller, consisting of 2-4 subunits with either SMARCA5 or SMARCA1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily A, members 1 and 5) as the catalytic subunit (Strohner et al., 2001). CHD complexes contain 1-10 subunits and have two tandemly arranged chromodomains on the N-terminus of the catalytic subunit, as reviewed in (Clapier & Cairns, 2009). In humans the CHD1 complex consists of just the catalytic subunit CHD1, whilst the NuRD complex is made up of many subunits (Delmas et al., 1993; Xue et al., 1998). INO80 complexes are also large and consist of over 10 subunits. They contain an ATPase domain that is split by a long insertion located in the middle of the domain and a helicase SANT (switching-defective protein 3, adaptor 2, nuclear receptor co-repressor, transcription factor IIIB)

associated domain, both of which are important for the recruitment of accessory subunits and complexes to INO80 (Clapier & Cairns, 2009). Chromatin remodelling complexes may be involved in a number of processes including gene regulation, DNA replication and chromatin assembly, and it is the combination of accessory and catalytic subunits that define the processes in which a complex may be involved (Park et al., 2006).

1.1.5 RNA mediated silencing

RNA interference (RNAi) refers to the degradation of messenger RNA (mRNA) product by double stranded RNA (dsSRNA) homologous in sequence to the RNA product (Fire et al., 1998). It can also broadly refer to the ability of small RNAs to modify chromatin and alter gene expression (Moazed, 2009). Long dsRNAs are processed into 22 nucleotide short interfering RNAs (siRNAs) by the RNase III ribonuclease, Dicer. The siRNA then binds Argonaut proteins within the RNA induced silencing complex (RISC), directing them to a complementary RNA and targeting them for degradation or transcriptional silencing (Elbashir et al., 2001; Meister et al., 2004; Moazed, 2009). Micro RNAs (miRNAs) are similar to siRNAs in that they act through the same RNA interference pathway. However, miRNAs arise from RNA transcripts that are folded back on themselves to form short hairpin loops (Bevilacqua & Blose, 2008).

A third class of small RNAs is PIWI (P-element induced wimpy testis) interacting RNAs (piRNAs). They are 26-30 nucleotides long, but unlike other small RNAs, have not been found to have a dsRNA precursor or require Dicer (Das et al., 2008; Houwing et al., 2007). In mice piRNAs have been shown to target transposons

in the male germline for *de novo* methylation resulting in their silencing (Kuramochi-Miyagawa et al., 2008). However, the exact mechanism by which this occurs in mammals is still to be deciphered.

Long non-coding RNAs (lncRNAs) are greater than 200 nucleotides long but do not encode proteins and have been found to play key roles in many processes such as chromatin remodelling, post transcriptional processing, and transcription, as reviewed in (Mercer et al., 2009). The lncRNA HOTAIR, Hox transcript antisense RNA, has the ability to direct Polycomb chromatin remodelling complexes to the HOXD locus, resulting in transcriptional silencing (Rinn et al., 2007). Whilst the lncRNAs NRON inhibits the transportation of NFAT to the nucleus via the binding of specific nucleocytoplasmic machinery (Willingham et al., 2005). While the functions of lncRNAs are many and diverse, few have been characterised to date, indeed future research may uncover more roles for lncRNAs.

1.1.6 Correlation between epigenetic marks

Recently it has been found epigenetic modifications do not act independently, rather they are able to work together and/ or cross regulate each other. Many studies have focused on the unique relationship between DNA methylation and histone methylation. The histone methyltransferase Ehmt2 (euchromatic histone lysine N-methyltransferase 2) has been found to be responsible for catalysing mono- and dimethylation of H3K9 in embryonic stem cells (ESC), and loss of Ehmt2 results in the loss of DNA methylation at CpG rich loci (Ikegami et al., 2007; Tachibana et al., 2008). In contrast, loss of the three DNA methyltransferases (DNMT1/DNMT3a/Dnmt3b) in ES cells has no impact upon H3K9 methylation with

cells retaining the ability to maintain an undifferentiated state (Tsumura et al., 2006). Whilst the mechanism through which many independent epigenetic modifiers has been studied in detail, there is still much to be learnt as to how the presence of different epigenetic marks influence one another.

1.2 EPIGENETIC PHENOMENA

The study of epigenetic regulation of gene expression has been investigated utilising different model systems including *Drosophila*, *C. elegans* and mammalian models. This review will focus on dosage compensation, parental imprinting and transgene silencing in mammals.

1.2.1 Dosage compensation

Female mammals have two X-chromosomes, whilst males have one X and one Y Chromosome. Early in embryogenesis, one of the female X-chromosomes becomes silenced leaving the other active (Lyon, 1961). Transcriptional output on this active X-chromosome is then increased to be equal to expression of autosomal genes, (Lin et al., 2007; Nguyen & Disteche, 2006). Interestingly, mammalian X-linked gene transcripts have been shown to be less susceptible to decay compared to autosomal gene transcripts and this may contribute to their greater abundance (Yin et al., 2009). Only cells that contain two or more X-chromosomes are able to undergo this process of X inactivation. The *Xic* locus (X-chromosome inactivation centre) located on the X-chromosome and which produces the non-coding RNA Xist, has been shown to play a role in ensuring only one X remains active (Kay et al., 1993; Penny et al., 1996). *Xist* RNA is expressed from the allele on the inactive X and

mediates gene repression in a chromosome wide manner (Brockdorff et al., 1992; Brown et al., 1992). Recent studies have identified further factors that are required for chromosomal repression by *Xist.* One example is Spen (Spen family transcriptional repressor), which is recruited by *Xist* and subsequently recruits Hdac3 followed by Polycomb repressive complex 1 accompanied by H3K27me (Chu et al., 2015; McHugh et al., 2015; Monfort et al., 2015). This is then followed by recruitment of Polycomb repressor complex 1 and Dnmt1 leading to the repression of the X chromosome (Cao et al., 2002).

1.2.2 Genomic imprinting

Whilst the majority of genes in the human genome are expressed equally when inherited from either parent, a small minority (~100 in mammals) are not, as reviewed in (Bartolomei & Ferguson-Smith, 2011). These genes undergo genomic imprinting, whereby expression is restricted to one of the two parental alleles in diploid cells (reviewed in Bartolomei et al., 2011). Studies investigating nuclear transplantation were the first to suggest genomic imprinting. These studies found embryos reconstructed from either two maternal or two paternal pronuclei failed to survive, when compared to embryos reconstructed from one maternal and one paternal pronuclei (McGrath & Solter, 1984; Surani et al., 1986). Interestingly, the small number of paternal imprinted genes identified are implicated in growth promotion, whilst maternal imprinted genes have functions in growth retardation. One example of this is the paternally expressed insulin like growth factor 2 gene, *Igf2*, and its maternally expressed receptor *Igf2r* (Barlow et al., 1991; Latham et al., 1994). Studies investigating knockout mice for both of these genes found mice null for paternally expressed *Igf2* were growth retarded, whilst those null for maternally

expressed *Igf2r* exhibited overgrowth (Lau et al., 1994). Many imprinted genes are found in clusters and are under the control of *cis*-acting loci called imprinting control regions which are differentially methylated in the germline (Kurukuti et al., 2006; Shin et al., 2008). This difference in methylation may be acquired either during gametogenesis or following fertilisation (Lewis & Reik, 2006).

1.2.3 Silencing of foreign DNA, including transgenes

Studies investigating the insertion of transgenes and foreign DNA into host genomes have found them to often be silenced via epigenetic mechanisms (Garrick et al., 1996; Martin & Whitelaw, 1996). Alterations in methylation at both the site of integration as well as at sites further away have been identified. Indeed the insertion of foreign DNA into the hamster genome has been shown to lead to marked alterations in certain cellular DNA segments, which is thought to be due to changes in chromatin structure (Remus et al., 1999). Transgenes have also been shown to be expressed in a variegated manner, whereby they are expressed in some, but not all, cells and this is thought to be influenced by their insertion site (Martin et al., 1996). Furthermore, studies in mice have shown differences in transgene copy number alter transgene expression (Garrick et al., 1998). Mice containing ~200 copies of a murine transgene showed a marked increase in silencing versus those carrying ~5-8 copies of the same transgene (Kearns et al., 2001).

1.3 EPIGENETIC REPROGRAMMING

Once the epigenetic state of an organism has been established, it is relatively stable and is maintained throughout the organisms' life (Morgan & Whitelaw, 2008). Epigenetic state is rarely passed on to the next generation, instead epigenetic reprogramming serves to erase epigenetic marks that have accumulated in the parent, so that the next generation has a 'clean slate' with which to initiate development (Dean et al., 2003; Hajkova et al., 2002). Genome-wide demethylation occurs immediately following fertilization of the zygote and also during establishment of the primordial germ cells, **Figure 1.2** (Gregoretti et al., 2004). In the zygote the genome contributed by each parent undergoes independent epigenetic remodelling. The paternal genome is actively demethylated before the first cell division is completed, and there has been recent evidence of paternal 5mC being actively converted to 5hmC and removed over subsequent cleavages (Inoue et al., 2011; Inoue & Zhang, 2011; Oswald et al., 2000), whilst the maternal genome is passively demethylated over successive cleavage divisions due to the exclusion of DNMT1 from the nucleus (Howlett & Reik, 1991). Daughter cells of the zygote therefore inherit a lowly methylated reprogrammed genome (Gregoretti et al., 2004). However, research has begun to look for evidence that this "clearing" of marks may not always be complete resulting in transgenerational epigenetic inheritance (Morgan & Whitelaw, 2008). This incomplete clearing is thought to provide a chance to adapt to environmental changes faster than genetic inheritance (Cerone et al., 1997).

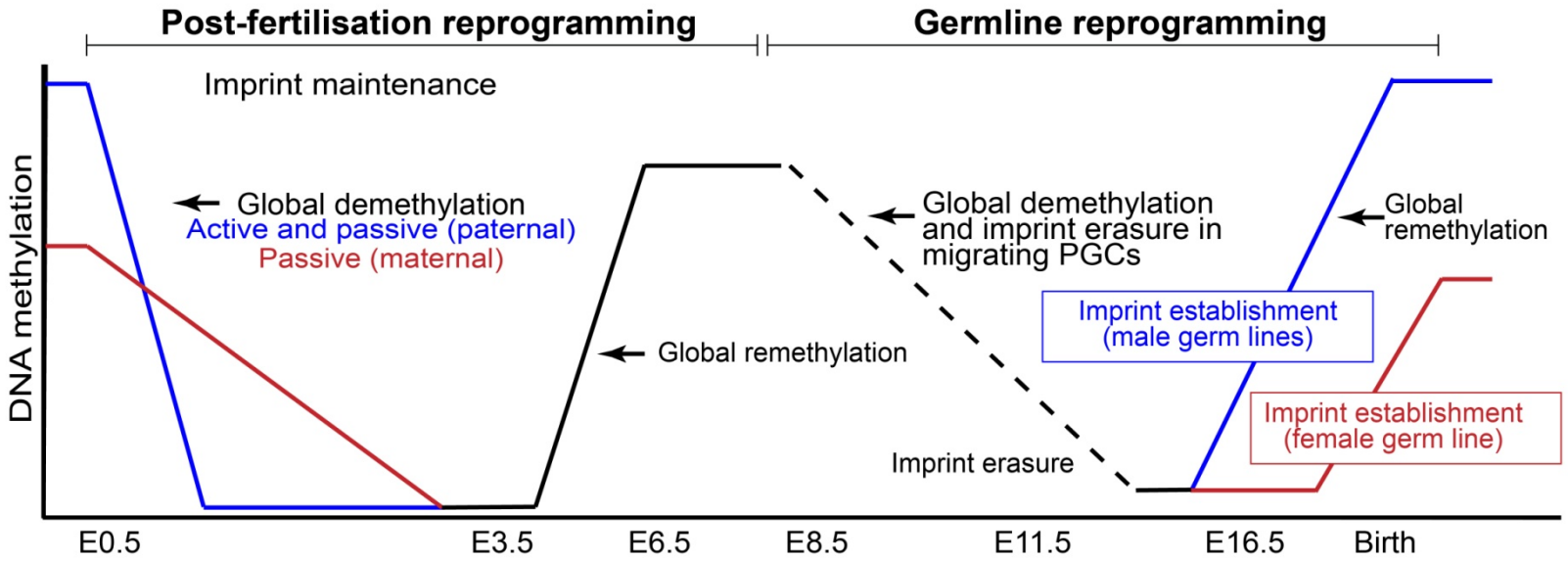


Figure 1.2: DNA methylation reprogramming in mammalian embryos.

Following fertilisation DNA methylation marks are erased in two waves. Firstly, the paternal genome (blue) undergoes rapid demethylation, followed by passive loss of methyl marks in the maternal genome (red). Re-establishment of DNA methylation is observed after the morula stage. The second wave occurs in primordial germ cells (PGCs) in which paternal and maternal marks are erased, followed by the re-establishment of imprinted genes in PGCs. Adapted from Cell, Vol 157, Issue 1, (Heard & Martienssen, 2014), Transgenerational Epigenetic Inheritance: Myths and Mechanisms, Pages No. 95-109, Copyright (2014), with permission from Elsevier; review. (**Appendix XII**).

1.4 TRANSGENERATIONAL INHERITANCE

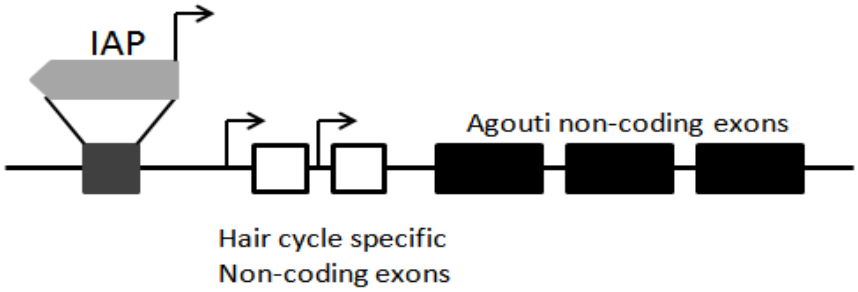
Transgenerational epigenetic effect refers to the transfer of non-genetic physical and behavioural information across generations. Transgenic epigenetic inheritance, however, refers to the transfer of chromosomal modifications to the next generation via the gametes, as reviewed in (Youngson & Whitelaw, 2008). To date there have been numerous studies investigating transgenerational epigenetic inheritance in plants, with many studies focusing on how changes to the environment can induce changes in the epigenome that are consequently inherited and expressed by progeny not exposed to the parental stressor (Norouzitallab et al., 2014; Zheng et al., 2013). Studies in C. elegans have also found epigenetic marks can be transmitted from on generation to the next. For example, loss of the histone demethylase lsd1 leads to an accumulation of histone methylation in the subsequent generation (Katz et al., 2009). Although there is some evidence for transgenerational epigenetic inheritance in mammals, the field remains highly controversial. In humans the study of transgenerational epigenetics is complex due to the inability to exclude genetic or societal reasons for the transmission of a phenotype across generations (Morgan & Whitelaw, 2008). This makes the utilisation of inbred genetically identical mouse colonies suitable models for studying transgenerational epigenetic inheritance.

1.4.1 A^{vy} metastable epiallele

Metastable epialleles, loci at which activity is dependent on the epigenetic state, provide molecular evidence for transgenerational epigenetic inheritance in the mouse (Youngson & Whitelaw, 2008). One such example is the *agouti viable yellow* (A^{vy}) allele, which can undergo stochastic gene silencing associated with variegation

similar to transgenes (Bultman et al., 1992; Rakyan et al., 2002). The agouti alleles are responsible for the production of black and yellow pigment in the hair follicles (Wolff et al., 1998). The A^{vy} metastable epiallele is the result of insertion of an intracisternal A particle (IAP) upstream of the agouti gene transcription start site, **Figure 1.3A** (Morgan et al., 1999). Mouse coat colour correlates with methylation state of a cryptic promoter in the 3' long terminal repeat (LTR) of the retrotransposon (Morgan et al., 1999). Isogenic mice carrying the A^{vy} allele can be yellow, mottled or pseudoagouti (dark brown), **Figure 1.3B** (Bultman et al., 1992). Hypomethylation of the cryptic promoter is associated with a yellow coat colour and hypermethylation is associated with silencing and a pseudoagouti coat. The presence of mottled mice illustrates that the allele is expressed in a variegated manner (Morgan et al., 1999).

A.



B.

Figure 1.3: Schematic diagram of A^{vy} allele and mice carrying the A^{vy}.

A. A^{vy} allele carries an IAP retrotransposon inserted upstream of agouti gene. **B.** Isogenic C57BL/6 A^{vy}/*mice* showing phenotypes ranging from completely yellow (left), through degrees of yellow/agouti mottling, to completely pseudoagouti (right). Adapted by Permission from Macmillan Publishers Ltd: Nature Genetics (Morgan et al., 1999), copyright (1999). (**Appendix XIII**).

1.4.2 Evidence for transgenerational inheritance in mice

Early studies of the A^{vy} locus found a transgenerational effect on coat colour when inherited from the dam but not the sire, as pseudoagouti dams produced more pseudoagouti offspring than mottled and yellow dams (Wolff et al., 1998). This was further supported by Morgan et al., 1999 who identified the cause being due to incomplete clearing of the epigenetic state in the female germline rather than being environmentally induced. More recent studies have utilised the A^{vy} allele to investigate the effect of methyl supplemented diet on transgenerational inheritance with conflicting results.

One study by Cropley et al., 2006, showed in utero exposure of A^{vy} mice to folate can shift coat colour of not only those mice, but also their offspring (Cropley et al., 2006). In this study pregnant pseudoagouti dams were fed a supplemented diet from E8.5 – E15.5. At which time the gametes of F1 offspring are being formed that will produce F2 offspring. The shift in coat colour of F2 offspring may have been influenced by the 'presence' of F2 gametes during F0 diet supplementation (Cropley et al., 2006). A more recent study by this group using the same supplementation method found this had no impact on methylation at the A^{vy} allele in either F1 or F2 offspring (Cropley et al., 2010). A separate study by Waterland et al., 2007, found no cumulative effect when investigating the effect of methyl rich diets across multiple generations (Waterland et al., 2007). This study was designed as a three generation cumulative exposure study, in which slightly yellow mottled dams (F0) were fed a supplemented diet throughout pregnancy and lactation and F1 and F2 females weaned onto the same diet as their mothers. The supplemented diet group consisted of more heavily mottled mice than the control group in each of the generations,

however no cumulative increase in the number of heavily mottled mice was observed. Waterland et al., 2007, results suggest diet induced changes are actively erased between each of the generations versus those of Cropley et al., 2006, which suggest environmental influences may allow for maintenance of epigenetic state in the germ line rather than the accumulation of new epigenetic marks (Waterland et al., 2007). Whether environmentally induced epigenetic changes can lead to transgenerational epigenetic inheritance is still poorly understood.

Whilst the majority of studies to date have investigated the retention or loss of epigenetic marks across generations, studies investigating the disruption of epigenetic reprogramming in the future may be more beneficial.

1.4.3 Evidence for transgenerational inheritance in humans

There are few studies investigating transgenerational epigenetic inheritance in humans. One study investigated the availability of food supply to grandparents with the mortality risk ratio of their grandchildren. This study of the Överkalix, northern Sweden cohort, found the grandfathers food supply was linked only to the mortality risk of grandsons, whilst the grandmothers was linked only with the risk ratio of granddaughters (Pembrey et al., 2005). However, a pitfall of this study, and most human studies investigating this phenomenon, is that transgenerational epigenetic inheritance is not the only possible explanation. It is complex to rule out genetic or societal reasons for observing the transmission of a particular phenotype across multiple generations in humans (Morgan & Whitelaw, 2008). In this context inbred mouse colonies offer a chance to study transgenerational epigenetic inheritance in a setting where each mouse is genetically identical.

1.5 DISEASES RESULTING FROM ABERRANT EPIGENETIC STATE

The mutation or inactivation of a number of epigenetic modifiers have been implicated in the development and/or progression of a wide variety of diseases in humans. The introduction of next generation sequencing has enabled the identification of causative mutations within epigenetic modifiers to be associated with diseases they had not previously been associated with.

1.5.1 Diseases related to aberrant control of imprinted regions

A number of disorders arise due to disruptions in imprinted genes/regions. Two well characterised but distinct disorders are Angelman syndrome (AS) and Prader-Willi syndrome (PWS), both of which arise from a deletion within an imprinted control region on chromosome 15 (Kishino et al., 1997; Ledbetter et al., 1981). Loss of the maternally expressed UBE3A, ubiquitin protein ligase E3A, a gene required for brain development results in AS (Kishino et al., 1997). Deletion of the maternal chromosome 15q11-13 region has been linked to a more severe phenotype than point mutations within this gene (Gentile et al., 2010). A number of paternally expressed genes within this same region have been linked with PWS. Unlike AS, where deficiency in a single gene has been linked to phenotype severity, it is thought deficiency in a number of genes in PWS results in a more severe phenotype (Angulo et al., 2015; Kanber et al., 2008). Interestingly, both PWS and AS can arise due to uniparental disomy, whereby the child may inherit two copies of the maternal allele (PWS) or paternal allele (AS) resulting in no active allele being present and manifestation of these disorders (Malcolm et al., 1991).

1.5.2 Disorders related to DNA methylation defects

Mutations in DNMT3B have been associated with Immunodeficiency, Centromeric region instability and Facial anomalies syndrome (Okano et al., 1999; Xu et al., 1999). Mutations in the C-terminal portion of both alleles of DNMT3B are commonly identified in patients with ICF (Hansen et al., 1999; Wijmenga et al., 2000). It has been predicted complete loss of DNMT3B is likely to be embryonic lethal, as observed in the mouse, and as such it is thought the mutations observed in DNMT3B cause a reduction in activity, rather than a loss of function (Jin et al., 2008; Okano et al., 1999). Another feature of ICF syndrome is hypomethylation of DNA satellites that are normally highly methylated, in lymphocytes this is combined with chromosomal instability suggesting DNMT3B function is of particular importance in this cell type (Kondo et al., 2000; Xu et al., 1999).

1.5.3 Disorders related to histone modifying enzymes

Loss of function of a number of histone modifying enzymes results in different disorders in humans. One example is Kabuki syndrome, in which patients display intellectual disability, distinct facial features and developmental delay (Niikawa et al., 1981). The majority of patients present with mutations in the histone methyltransferase KMT2D, lysine (K)-specific methyltransferase 2D, while a small number present with mutations in KDM6A, lysine (K)-specific demethylase 6A (Hannibal et al., 2011; Lederer et al., 2012; Ng et al., 2010). Another group of patients displaying classical Kabuki phenotype were found to have no mutations in either of these genes. As both KDM6A and KMT2D are associated with the same large protein complex, mutations in other proteins within this complex may be suitable candidate genes (Van Laarhoven et al., 2015). Interestingly, KDM6A

functions in removing the silencing histone mark H3K27me3, while KMT2D functions in removing the activating histone mark H3K4me3 suggesting the maintenance of histone methylation is critical for normal development. (Hübner et al., 2013).

1.5.4 Disorders related to chromatin remodelers

Chromatin remodelers have been shown to be critical for normal development and mutations in different complex subunits can lead to disease. One disorder is Coffin-Siris syndrome characterised by mental retardation, developmental delay and phalange defects (Santen et al., 2012; Tsurusaki et al., 2012). The majority of Coffin-Siris patients carry one or more mutations in both catalytic and non-catalytic subunits of the SWI/SNF complex (Santen et al., 2012; Tsurusaki et al., 2012). Regardless of the different mutations identified patients with Coffin Siris syndrome present with similar phenotypes, suggesting alterations to any of the SWI/SNF subunits can result in aberrant gene regulation (Tsurusaki et al., 2012). This is further supported by studies investigating Nicolaides–Baraitser syndrome, which is caused by mutations within the subunit SMARCA2 that are not found in patients with Coffin-Siris syndrome (Sousa et al., 2009).

1.6 A SCREEN FOR EPIGENETIC MODIFIERS IN THE MOUSE

Screens investigating genes that affect variegated expression have been undertaken in *Drosophilia* and the mouse (Schotta et al., 2003). Past screens in the mouse were utilised to identify genes implicated in parental imprinting and X-inactivation (Percec et al., 2002; Tsai et al., 2002). The Whitelaw laboratory was the first to

design a screen for identifying genes involved in establishing and maintaining epigenetic state in the mouse (Blewitt et al., 2005).

1.6.1 Modifiers of murine metastable epialleles

An N-ethyl-N-nitroso urea (ENU) random mutagenesis screen was established by Professor Whitelaw and colleagues to identify novel epigenetic modifiers in the mouse (Blewitt et al., 2005). The mice used in this study carry a multi-copy green fluorescent protein (GFP) transgene array linked to an erythroid specific α-globin promoter and a HS40 enhancer (Preis et al., 2003). Multi-copy transgenes have been shown to undergo silencing due to the formation of repressive chromatin environments via epigenetic mechanisms that are locus specific (Garrick et al., 1998). This results in variegated expression of GFP in the red blood cells, i.e. only 55% of red blood cells express GFP, as it is silenced in 45% of cells, **Figure 1.4**. Alleles that variegate in this way are known as metastable epialleles (Rakyan et al., 2002). A male mouse was injected with the chemical mutagen ENU and mated to females carrying the transgene (Blewitt et al., 2005). Mutation of an epigenetic modifier is expected to lead to a shift in the percentage of red blood cells expressing GFP, and offspring in which the percentage of GFP expression has been significantly altered are identified and named a *MommeD* (Modifier of murine metastable epiallele Dominant) (Blewitt et al., 2005).

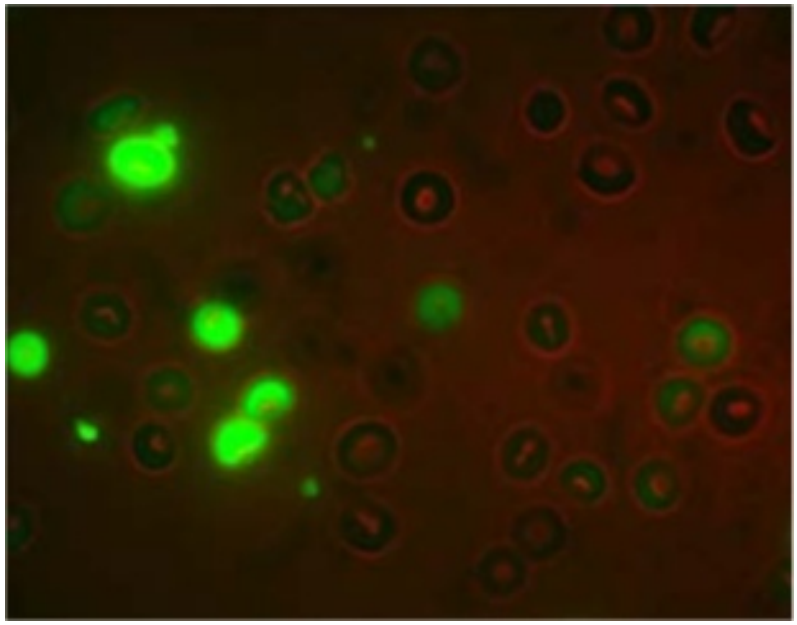


Figure 1.4: Variegated expression of GFP in erythrocytes.

GFP is expressed in 55% of erythrocytes in wild-type mice, while it is silenced in the remaining 45%. Adapted from Daxinger et al., 2013.

Over 60 mutant mouse lines were identified from the *Momme* screen. Causative genes included known epigenetic modifiers such as *Dnmt1*, validating the screen (Ashe et al., 2008; Chong et al., 2007; Daxinger et al., 2013). Novel epigenetic modifiers have also been identified, for example structural maintenance of chromosomes hinge domain containing 1 (*Smchd1*), encodes a formerly uncharacterised protein that was found to play a critical role in X inactivation (Blewitt et al., 2008). *Rlf* is a second novel epigenetic modifier identified from the *Momme* screen (Daxinger et al., 2013).

1.6.2 The first mouse mutants of Rearranged L-Myc Fusion

Three mouse lines from the screen were found to have mutations in *Rlf*, **Figure 1.5A**. *MommeD8* has a mutation changing a cysteine to a phenylalanine in the last predicted zinc finger in exon 8 of *Rlf* (Daxinger et al., 2013). *MommeD28* was found to have an adenine → guanine mutation at the 3′ splice site of intron 4, and a point mutation changing a cysteine to a stop codon was detected in *MommeD34* (Daxinger et al., 2013). A decrease in GFP expression was observed in all three lines compared to wild-type mice, showing *Rlf* to be an enhancer of variegation (Daxinger et al., 2013). Western blotting revealed a ~280 kDa band in protein lysates made from E12.5 wild-type embryo heads, **Figure 1.5B**. The band was absent in $Rlf^{MommeD28/MommeD28}$ and $Rlf^{MommeD34/MommeD34}$ heads, suggesting presence of a null allele, and reduced in $Rlf^{MommeD8/MommeD8}$ heads, suggesting a hypomorphic allele (Daxinger et al., 2013). Currently little is known about how *Rlf* acts as an epigenetic modifier.

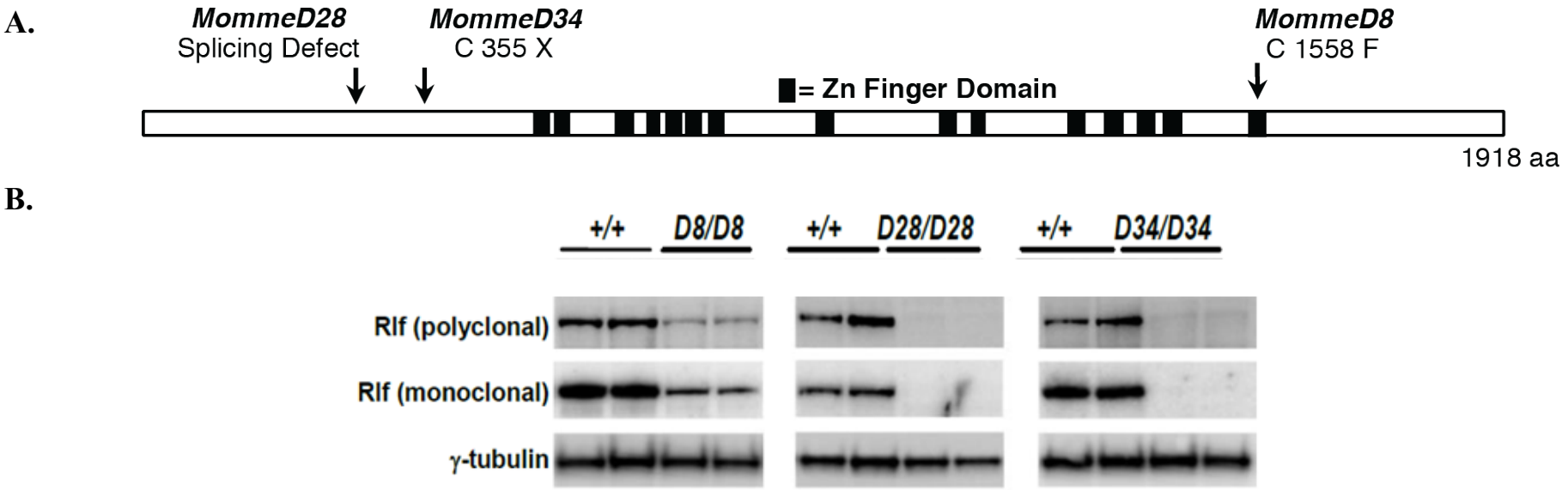


Figure 1.5: Schematic diagram of *Momme* mutations in *Rlf,* and expression of Rlf in *Momme* mutants.

A. Location of ENU point mutations in Rlf for the *MommeD8*, *MommeD28* and *MommeD34* mouse lines. **B.** Western blotting of Rlf in embryo head lysates from E12.5 $Rlf^{+/+}$, $Rlf^{MommeD28/MommeD28}$, $Rlf^{D34/34}$ (both no protein expression) and $Rlf^{MommeD8/MommeD8}$ embryos (some protein expression), using commercially available polyclonal Rlf antibody (Abcam, Ab115011) and monoclonal Rlf antibody (Abnova, M05, clone2G2) (Daxinger et al., 2013).

1.6.3 *Rlf* is a novel epigenetic modifier

Alterations in DNA methylation at the transgene were investigated utilising bisulphite sequencing of wild-type and *MommeD8*, *MommeD28* and *MommeD34* homozygous mutant mice. Bisulphite sequencing allows the methylation status of single cytosines to be determined, since addition of bisulphite converts unmethylated cytosines to uracil, whilst methylated cytosines remain unchanged (Clark et al., 1994; Frommer et al., 1992; Hyatsu et al., 1970). Analysis of the enhancer of the GFP transgene revealed an increase in methylation in all three *Rlf* mutant lines compared to wild-type littermates (Daxinger et al., 2013).

Haploinsufficiency for modifiers of epigenetic reprogramming have been shown to alter the ratio of coat colours in mice carrying the A^{vy} allele (Blewitt et al., 2005; Chong et al., 2007; Gaudet et al., 2004). The effect of haploinsufficiency for *Rlf* on the endogenous A^{vy} allele has also been investigated. *MommeD8* heterozygous mice were mated to pseudoagouti mice carrying the A^{vy} allele and the offspring scored for coat colour. It was found that offspring that inherited the *MommeD8* allele were more likely to be pseudoagouti than their wild-type littermates, indicating that haploinsufficiency for *Rlf* increases the probability of silencing at the A^{vy} locus (Daxinger et al., 2013). This result is consistent with the initial observation that *Rlf* acts as an enhancer of variegation.

Taken together the above experiments provide substantial evidence for the involvement of *Rlf* in regulating epigenetic state. However, the mechanism via which this occurs is still unclear and no previous studies have investigated the consequences of loss of *Rlf* in the mouse.

1.7 REARRANGED L-MYC FUSION

RLF was first identified as part of a gene fusion in human small cell lung cancer (SCLC), and more recently has been identified as a putative binding partner with other chromatin interacting proteins (Eberl et al., 2013; Makela et al., 1991b; Nozawa et al., 2010). These key papers will be summarised here.

1.7.1 Gene fusion in small cell lung cancer

A recurrent *RLF-MYCL1* (v-Myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog) gene fusion has been detected in multiple studies in human SCLC cell lines and primary tumours (Makela et al., 1995; Makela et al., 1991a; Makela et al., 1992; Rudin et al., 2012). This fusion protein consists of the first 79 amino acids of RLF joined to three MYCL1 noncoding amino acids and the second and third exons of MYCL1, although this was found to differ slightly between cell lines (Makela et al., 1991b). Although previous studies have investigated the function of the RLF-MYCL1 fusion protein, the function of RLF itself has not been investigated.

1.7.2 Interaction of RLF with chromatin

Recently, two independent studies using human cell lines and mouse tissue have utilised proteomics to identify proteins that bind to H3K9me3. Whilst human and mouse Rlf were not found to directly interact with this histone mark they do bind to HP1α (Heterochromatin protein 1α), a protein that localizes to heterochromatin and is thought to be critical for initiating and maintaining heterochromatin

conformation (Cheung & Lau, 2005; Nozawa et al., 2010). HP1α contains a chromodomain (CD) and a chromoshadow domain (CSD) separated by a hinge region (Tahiliani et al., 2009). The CD recognizes methylated H3K9, while the CSD is responsible for dimerisation of HP1α creating a hydrophobic surface that binds to HP1α binding proteins containing a PxVxL motif (Eden et al., 1998; Hashimshony et al., 2003; Tsukada et al., 2006).

The first study found human RLF bound to a full-length flag tagged HP1α, as well as two mutants with deletions in the CSD, and histone binding domains. A mutant carrying a single amino acid substitution in HP1α's PxVxL binding domain was found to not bind RLF (Nozawa et al., 2010). Preliminary analysis of both human and mouse Rlf protein coding sequence reveals a PxVxL (PVVVL) motif approximately 30 amino acids from the carboxyl end (L. Bourke, unpublished). This motif may interact directly with HP1α, however this has not been tested. The second study linked mouse Rlf with HP1α in histone peptide H3K9me3 pull-downs with quantitative mass spectrometry in mouse brain and liver tissue (Eberl et al., 2013). While Rlf was identified as a putative interacting protein in both of these proteomics studies, no further validation of the Rlf interaction was performed.

1.7.3 HP1α interacting proteins

Analysis of the primary amino acid sequence of *Rlf* revealed 15 putative Cys_2His_2 zinc fingers domains, and also a putative PxVxL motif (Makela et al., 1995) (Bourke, unpublished). Proteins containing either a Cys_2His_2 zinc finger domain or PxVxL motif have been found to bind to HP1α. In both studies that identified Rlf as

a potential binding partner of HP1α, a large number of zinc finger containing proteins were also pulled down. Some of these may destabilise HP1α binding or they may enable the localisation of HP1α by facilitating heterochromatin formation. One example of this is ZFP462 (Zinc finger protein 462) where loss of this protein disrupted heterochromatin resulting in cell death due to HP1α mislocalisation (Tamaru & Selker, 2001). Another protein whose association with HP1α has been studied in depth is POGZ (Pogo transposable element with ZNF domain). POGZ was found to bind to HP1α through a Cys_2His_2 zinc finger like motif. This binding region competes with HP1α binding partners containing PxVxL sequence and destabilises the interaction of HP1α with chromatin (Nozawa et al., 2010). However both of these mechanisms of action suggest HP1α binding partners are associated with gene repression, this is in disagreement with the *Momme* screen where Rlf has been linked with gene activation.

Although small amounts of information about the role of Rlf can be gathered from these global approaches. More direct studies into the function of Rlf are required to gain a better understanding.

1.8 AIMS OF THIS THESIS

Currently little is known about how Rlf acts as an epigenetic modifier. This project will utilise the first mouse mutants of *Rlf* to understand the effects of loss of Rlf on phenotype, and to gain a clearer understanding as to the molecular consequences of loss of Rlf in the mouse.

The hypothesis of this thesis is:

Rearranged L-Myc Fusion, *Rlf*, is a novel epigenetic modifier of critical importance in development and disease in the mouse.

The major aims of this thesis are:

Aim 1: To analyse the effect of loss of Rlf in E14.5 fetal mouse liver

Previous work by the Whitelaw laboratory showed loss of Rlf results in an increase in methylation at the GFP transgene (Daxinger et al., 2013). The fetal liver was chosen for this study as at E14.5 it is the hematopoietic organ of the embryo and the *Momme* screen assesses transgene variegation in red blood cells. This thesis will investigate how loss of Rlf affects methylation across the genome, analyse the transcriptome to identify genes that are directly and indirectly regulated by Rlf, and identify putative Rlf binding sites in the fetal liver.

Aim 2: To determine the consequence of loss of Rlf on mouse development

Previous studies by the Whitelaw laboratory have shown mice homozygous null for *Rlf* are not viable at three weeks of age (Daxinger et al., 2013), and heterozygous null *Rlf* offspring show a reduction in weight compared to wild-type

mice (Ashe et al., 2008; Daxinger et al., 2013). However no further investigations into the loss of Rlf on phenotype have been undertaken. I have chosen to investigate the impact of loss of Rlf on embryonic development, and also the impact of reduced Rlf expression on ageing.

Aim 3: To investigate the consequence of reduced Rlf expression in adult mice

Epigenetic reprogramming resets the epigenetic marks accumulated in the parent so that the next generation has a 'clean slate' to initiate development from (Dean et al., 2003; Hajkova et al., 2002). Incomplete clearing of these epigenetic marks is referred to as transgenerational epigenetic inheritance (Morgan & Whitelaw, 2008). In this thesis I will assess whether reduced Rlf expression alters epigenetic reprogramming across multiple generations by assessing effects on litter size, weight, and methylation of the GFP transgene and at an endogenous locus.

Chapter 2: Materials and Methods

2.1 MATERIALS

2.1.1 Chemicals and reagents

Following is a list of all the major chemicals and reagents, in alphabetical order, used while conducting this project, along with the suppliers they were purchased from.

All chemicals are molecular biology grade, unless otherwise stated.

5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, **Applichem**

Agar (bacteriological grade), **Sigma-Aldrich**

Agarose (for electrophoresis), **Sigma-Aldrich**

Albumin from bovine serum lyophilised powder, **Sigma-Aldrich**

Ampicillin salt, **Progen Industries**

Chloroform:Isoamylalcohol 24:1, **Sigma-Aldrich**

Deoxyribosenucleoside triphosphates

(dATP, dCTP, dGTP, dTTP), **Bioline**

Diethylpyrocarbonate, **Sigma-Aldrich**

Dithiothreitol (DTT), **Bio-Rad**

Ethanol, **Ajax Finechem**

Ethidium Bromide, **Sigma-Aldrich**

Ethylenediaminetetra acetic acid disodium dehydrate (EDTA), **Sigma-Aldrich**

Glacial Acetic Acid, **Ajax Finechem**

Glycerol, **ChemSupply**

Glycine, **Ajax Finechem**

Glycogen, Life Technologies/**Invitrogen**

Isoflurane, **Abbott Australasia**

Isopropanol, **Ajax Finechem**

Isopropyl β-D-1-thiogalactopyranoside, **Applichem**

Methanol, **Ajax Finechem**

Osmosol, **POCD Healthcare**

Paraformaldehyde, **Sigma-Aldrich**

Phenol (pH 7.9), **Amresco**

Phenylmethylsulfonyl fluoride, **Sigma-Aldrich**

Phosphate buffered saline tablets, **Sigma-Aldrich**

Protease Inhibitor Cocktail tablets, **Roche**

Sodium chloride, **Sigma-Aldrich**

Sodium dodecyl sulphate, **Sigma-Aldrich**

Tris-HCl, **Sigma-Aldrich**

Tween-20, **Sigma-Aldrich**

Urea, **Sigma-Aldrich**

β-mercaptoethanol, **Sigma-Aldrich**

Dimethylformamide, **Ajax Finechem**

2.1.2 Enzymes

All enzymes were used as per their manufacturer's instructions.

HpyCH4Vrestriction endonuclease, **New England Biolabs Inc.**

Proteinase K from *Tritirachium album,* **Astral Scientific.**

RNase A (from bovine pancreas), **Roche Molecular Biochemicals**

2.1.3 Buffers and solutions

All solutions and buffers, alongside of their formulas and concentrations, are listed below. Milli-Q water was used as a diluents for all buffers and solutions.

10 X PBS	1.45 M NaCl, 10 mM Na_2HPO_4, 3 mM KH_2PO_4
10 x Transfer Buffer	10 mM Tris, 100 mM glycine, 10% methanol, 0.005% SDS
100mM PMSF	Made in 100% isopropanol
4% PFA solution	20 g PFA, in 1 X PBS, pH adjusted with 1 M NaOH to 7.0
4 X Loading dye	12.5 ml 4 X TrisCl/SDS (pH 6.8), 10 ml gycerol, 2 g SDS, 1.55 g DTT
4 X TrisCl/SDS (pH 6.8)	0.5 M TrisCl, 0.4% SDS, pH adjusted with 1 M HCl to 6.8
5 X Running Buffer	25 mM Tris, 0.1% SDS, 200 mM glycine
50 X TAE	0.2 M Tris-acetate, 5 mM EDTA (pH 8.0), 10% Glacial acetic acid
Blocking Solution	5% Skim milk, 1% BSA, PBS-Tween20
DNA lysis buffer	50 mM Tris, 10 mM EDTA, 100 mM NaCl, 1% SDS
IPTG	100 mU (23.8 mg/ml IPTG powder), diluted in H_2O
PBS-Tween 20	PBS, plus 0.1% (v/v) Tween20
Urea lysis buffer	8 M urea, 1/10 vol glycerol, 1/20 vol 20% SDS,1/200 vol 1 M DTT, 1/100 vol 1 M Tris (pH 6.8)
X-gal	20 mg/ml X-gal powder in DMF
TFB I	30 mM $KC_2H_3O_2$, 50 mM $MnCl_2$, 100 mM KCl, 10 mM $CaCl_2$, 15% (v/v) glycerol titrated to pH 5.8 with 0.2 M acetic acid
TFB II	10 mM MOPS (pH 7.0) 75 mM $CaCl_2$, 10 mM KCl, 15% (w/v) glycerol

2.1.4 Bacterial strains and microbial media

The bacterial strain used for all plasmid manipulations was *Escherichia coli* DH5α [supE44 DlacU169 (f80lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1], Bethesda Research Laboratories (Gaithersburg, MD, USA).

Bacteriological culture media was dissolved in Milli-Q water and sterilised by autoclaving.

LB broth 10 g/L Bacto® tryptone peptone digest (Amyl Media Pty. Ltd.), 5 g/L Bacto® yeast extract (Amyl Media Pty. Ltd.) 10 g/L NaCl. The medium was stored at 4 °C and ampicillin (from a filter sterilised stock) was added to a concentration of 100 μg/ml prior to inoculation.

LB agar LB broth supplemented with 15 g/L bacteriological agar (Amyl Media Pty. Ltd.). After the medium had cooled to 55 °C, ampicillin (from a filter sterilised stock) was added to a concentration of 100 μg/ml.

SOB broth 20 g/L Bacto® tryptone peptone digest (Amyl Pty. Ltd.), 5 g/L Bacto® yeast extract (Amyl Pty. Ltd.),10 mM NaCl, 10 mM KCl, 10 mM $MgSO_4$, 10 mM$MgCl_2$. SOB medium was prepared by adding tryptone, yeast extract and NaCl to Milli-Q water, adjusting the pH to 7.0, autoclaving, and adding filter sterilised stocks of $_{MgSO4}$ and $MgCl_2$.

2.1.5 Commercial kits

All commercial kits, and their suppliers, used throughout this project are listed below in alphabetical order.

AllPrep DNA/RNA Mini and Micro Kit, Qiagen

Big Dye Terminator kit v3.1, Applied Biosystems

Bradfords Reagent, Thermo Fisher Scientific Inc.

Clarity™ Western ECL blotting substrate, Bio-Rad

EpiTect Bisulphite Kit, Qiagen

GeneRuler™ DNA ladder mix, Thermo Fisher Scientific Inc.

iScript™ Reverse Transcription Supermix for qRTPCR, Bio-Rad

Mango*Taq*™ DNA Polymerase, which contains Taq DNA polymerase from *Thermus aquaticus,* Bioline

NucleoSpin Gel and PCR Clean-up, Macherey-Nagel

pGEM®-T Easy Vector, Promega

Platinum® SYBR® Green qPCR SuperMix-UDG, Invitrogen

Precision Plus Protein™ Kaleidoscope™ Standard, Bio-Rad

2.1.6 Primer sequences

All primers used in this project, along with PCR conditions, oligo sequences and electrophoresis protocols are shown in **Appendix I**. All primers were purchased from Integrated DNA Technologies.

2.1.7 Mouse strains and housing conditions

Ethical approval for the use of transgenic mice in this project was provided by the Animal Ethics Committee at QIMR Berghofer Medical Research Institute (QIMRB) (P2037 A081207) and QUT Animal Ethics Committee (1200000388).

The following *Mus musculus* strains were used:

1. Line3 transgenic line, produced in FVB/NJ mice by Dr J. Priest in the Whitelaw laboratory (Preis et al., 2003).
2. FVB/NJ mice carrying the $Rlf^{MommeD8}$ allele from Prof E. Whitelaw of the Whitelaw laboratory.
3. FVB/NJ mice carrying the $Rlf^{MommeD28}$ allele from Prof E. Whitelaw of the Whitelaw laboratory.
4. FVB/NJ mice carrying the $Rlf^{MommeD34}$ allele from Prof E. Whitelaw of the Whitelaw laboratory.
5. FVB/NJ mice carrying the $Rlf^{MommeD8}$ and $Rlf^{MommeD34}$ alleles from Prof E. Whitelaw of the Whitelaw laboratory.

All mice were housed at a temperature between 21 and 23 °C with a light/dark cycle of 12 hours. The mice had unlimited access to food and water. Male and female mice were placed in separate cages, unless they were being used for breeding experiments. Mice were weaned 21 days after birth.

2.2 GENERAL METHODS

2.2.1 Isolation of genomic DNA

Phenol: Chloroform method

Genomic DNA was isolated from adult tail, embryonic tail/limb or yolk sac tissue. Adult tail biopsies were collected from mice anesthetised with isoflurane (Abbott Australasia) using an anesthetising machine until sedated and a 0.2 cm section of tail removed. Embryo and yolk sac tissue were obtained by sacrificing pregnant dams at specified time-points after the observation of a copulation plug, noted as E0.5, and dissecting embryos in PBS. Proteolytic digestion was achieved by the addition of 300 µl of DNA lysis buffer containing proteinase K (Astral Scientific) at 1 mg/ml and incubating at 55 °C overnight. Organic extractions were sequentially performed in equal volumes of phenol (Amresco) and chloroform:isoamyl alcohol (Sigma) (total volume of 300 µl each), by vigorous shaking at room temperature followed by centrifugation (1,300 rpm, RT) for 5 min each. The organic phase was discarded after each extraction. High molecular weight chromosomal DNA was precipitated by adding 2 volumes of absolute ethanol (600 µl) to the final aqueous phase and mixed by inversion. The sample was centrifuged (1,300 rpm, 5 min RT), liquid removed and precipitated DNA washed with 70% ethanol. The ethanol was removed and the DNA pellet left to air-dry. DNA was then resuspended in ~100 µl of Milli-Q water. Concentration of the DNA extracted was determined spectrophotometrically with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific), assuming that an A260 nm of 1.0 corresponds to 50 µg/ml of dsDNA (Sambrook & Russell, 2001).

2.2.2 Genotyping by Polymerase Chain Reaction (PCR)

MommeD8, Momme28 and MommeD34 genotyping

Genotyping for the *MommeD8, MommeD28* and *MommeD34* alleles was performed using PCR followed by restriction enzyme digest (*MommeD8)* or Sanger sequencing (*MommeD28, MommeD34).* PCR was carried out in a Bio-Rad S1000 Thermal Cycler. Each reaction consisted of approximately 20 ng/µl of template gDNA, 0.1 U of Mango*Taq*™ DNA polymerase (Bioline), 200 µM of dNTPs (dATP, dCTP, dGTP, dTTP), 1.5 mM of $MgCl_2$, 500 nM each of forward and reverse primer for each allele, and 5X Mango*Taq*™ PCR reaction buffer. Primer sequences for each of the *Momme* alleles are shown in **Table 2.1**. The PCR reaction for all *Momme* alleles was 2 min of initial denaturation at 94 °C, then 35 cycles of denaturation (94 °C, 30 s), annealing (56 °C, 30 s) and extension (72 °C, 1 min), with final extension for 5 min at 72 °C. *MommeD8* allele primers produced a 500 bp product, *MommeD28* primers a 400 bp product and *MommeD34* primers a 500 bp product. These products were resolved on a 1.5% TAE (w/v) agarose gel.

2.2.3 Gel electrophoresis and purification of DNA fragments

PCR reactions were subject to gel electrophoresis on a 1.5% agarose (Sigma) gel in 1X TAE. Agarose gels contained ethidium bromide at 0.5 µg/ml. From each sample, 3-5 µl of PCR product was loaded on agarose gels, adjacent to one lane containing the DNA ladder GeneRuler™ marker (Thermo Fisher Scientific). Electrophoresis was carried out at 1-8 V/cm until the dye front had migrated an appropriate distance. Visualisation of agarose gels was performed by ultraviolet transillumination and photographed using a Bio-Rad Gel Imaging System. Following visualisation of a band at the appropriate size PCR products from *MommeD28* and

MommeD34 mice were purified using the NucleoSpin Gel and PCR Cleanup Kit according to manufacturer's instructions.

2.2.4 Restriction endonuclease digestion of DNA

Restriction endonuclease digestion of *MommeD8* PCR products was undertaken using HpyCH4V (New England Biolabs Inc). PCR product was digested with 0.1 U of restriction enzyme and 1X NEB Buffer 4 at 37 °C overnight and digested products resolved on a 1.5% TAE (w/v) agarose gel. $Rlf^{+/+}$ mice produced two bands of 300 and 200 bp each, $Rlf^{MommeD8/+}$ mice three bands of 500, 300 and 200 bp, and $Rlf^{MommeD8/MommeD8}$ mice one band at 500 bp.

2.2.5 PCR sequencing

Recovered PCR products were cleaned using the NucleoSpin Gel and PCR Cleanup kit (Macherey-Nagel) according to manufacturer's instructions and resuspended in 20 µl of Elution buffer (from Kit). Concentration of recovered product was determined with a NanoDrop 1000 Spectrophotometer. Approximately 1-2 ng per 100 bp PCR product was used in ABI Big Dye 3.1 sequencing reactions with 800 nM of forward or reverse primer, 0.6 µl Dye Terminator and 4.2 µl 5X reaction buffer in a final volume of 12 µl. Sequencing reactions underwent thermo cycling in a Bio-Rad S1000 Thermo Cycler with the following conditions: 96 °C for 1 in, followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s, 60 °C for 3min. Sequencing products were then cleaned by adding 72 µl 70% isopropanol followed by vortexing and incubating for 15 min at RT. Samples were then centrifuged for 30 min at maximum speed at 4 °C, supernatant removed and pellet rinsed in 200 – 300 µl 70% isopropanol and re-spun for 5-10 min. Again all liquid was removed and

samples air-dried at room temperature, before precipitated pellet was sent to QIMRB Sequencing Facility for capillary sequencing.

2.2.6 Western blotting

Protein extraction

Protein was extracted from tissue using urea lysis buffer. Adult tissue was homogenised using a handheld homogeniser (IKA) and embryonic tissue by passing tissue through a 19-guage needle with syringe until single cell was suspension observed. Samples were sonicated (3 x 10 s) using a Branson Sonifier 450A (Consonic), centrifuged at 13,000 g for 10 min at 4 °C and the supernatant collected for Western blot analysis. Protein concentration was measured using BCA protein assay reagent (Thermo Fisher Scientific), using 5 µl of sample and 200 µl of reagent as per the manufacturer's instructions. Protein yield was quantified by comparing to a standard curve of 1 – 10 µg/µl BSA. All samples were read at 595 nm on a VersaMax microplate reader (Molecular Devices).

SDS-PAGE

Protein (25 – 30 µg) was heated to 98 °C for 10 min in urea lysis buffer and 6X loading dye. Samples were then loaded onto a 4 – 15% Mini-PROTEAN® TGX™ Precast gel (Bio-Rad), and electrophoresed for approximately 120 min at 130 V in 1X Running buffer. The Precision Plus Protein™ Kaleidoscope™ pre-stained protein standards molecular weight marker was used to enable protein size estimation.

Western blot protein transfer

PVDF membrane (Merck Millipore) was activated in 100% methanol for 15 s and immersed in Milli-Q water for at least 2 min. The foam pads and filter paper were also pre-soaked in 1X Transfer buffer for 5 min. Separated proteins were then transferred to a membrane for 1 – 2 hours at 130 V and 4 °C in 1x Transfer buffer.

Membrane blocking and antibody incubations

The membrane was blocked for 30 min in blocking solution, and incubated with primary antibody overnight at 4 °C; antibodies were used at 1:100 – 1:1000 dilutions. The membrane was washed 3 times in PBS-Tween for 5 min each. Followed by an hour incubation at room temperature with secondary horseradish peroxidise antibody diluted in blocking solution (1:5000). The membrane was then washed 5 times in PBS-Tween, 5 min each. The following primary antibodies were used: RLF (Ab115011, Abcam), custom designed Rlf (Ab1, Ab2, Ab3, GenScript) and γ-tubulin (SAB4503045, Sigma-Aldrich). Secondary antibodies used were polyclonal goat anti-rabbit immunoglobulins/HRP (Dako) and polyclonal goat anti-mouse immunoglobulins/HRP (Dako).

Chemiluminescent detection

Clarity™ Western ECL substrate (Bio-Rad) was mixed according to manufacturers' instructions. Excess wash buffer was drained from membrane and ECL detection substrate pipetted onto the membrane surface and incubated for 5 min. The membrane was then visualised using MF-ChemiBis 3.2 (DKSH) or ImageQuant LAS 500 (GE Healthcare).

2.2.7 Quantitative reverse transcription Real-time PCR

RNA extraction

Total RNA was extracted using either an AllPrep Mini kit (Qiagen) (adult tissue, fetal liver) or AllPrep Micro kit (Qiagen) (fetal hearts) according to manufacturer's instructions. Fresh tissue (< 30 mg for AllPrep Mini kit or < 5 mg AllPrep Micro kit) was homogenised in RLT Plus buffer supplemented with β-mercaptoethanol using a needle and syringe. RNA was eluted in 30 µl of RNase free water (Sigma), and the eluate re-run through the column a second time to increase final RNA concentration. Samples were stored at -80 °C until required.

cDNA synthesis

cDNA was synthesised from total RNA using iScript™ Reverse Transcription Supermix for qRT-PCR (Bio-Rad), according to manufacturers' instructions. Total RNA concentration was quantified using a NanoDrop 1000 Spectrophotometer and 1 µg of total RNA was mixed with 4 µl of iScript RT Supermix and RNase free water to a total reaction volume of 20 µl. Samples were incubated in a Bio-Rad S1000 Thermal Cycler for 5 min at 25 °C, 30 min at 42 °C and 5 min at 85 °C. cDNA was then stored at -20 °C prior to use.

qRT-PCR

qRT-PCR was performed with Platinum® SYBR® Green qPCR SuperMix UDG (Life Technologies), following manufacturers' instructions. SYBR Green mix (including 3mM $MgCl_2$), 200 nM forward and reverse primers (**Appendix I**) and 10 ng cDNA were mixed together. The reaction was run on a Corbett Rotorgene™ 6000 using the following conditions: 95 °C for 10 min, followed by 40 cycles of 95 °C

for 15 s and 60 °C for 30 s, a melt curve of 95 °C for 15 s and 60 °C for 30 s, and a ramp of 62 °C, to 95 °C, adding 1 °C per 5 s. Primers were designed to span exon/intron boundaries to eliminate the risk of gDNA contamination. All samples were run in triplicate alongside no template and β-actin controls for each sample in the same reaction. Data from each run were analysed separately, and triplicates of less than 0.5 cycle thresholds (Ct) apart were tolerated. Transcript relative levels were calculated using the comparative Ct (ΔΔCt) method (Livak & Schmittgen, 2001):

1. Calculate the first ΔCt: gene of interest (Ct) minus reference gene (Ct) for each sample.
2. Calculate the ΔΔCt: subtract each ΔCt with the biggest ΔCt (will become calibrator from step 1.
3. Calculate the relative copy number by $2^{-(\Delta\Delta Ct)}$. One value becomes '1' (calibrator).

2.3 EMBRYO DISSECTIONS AND HISTOLOGY

Timed matings were performed by pairing male and female mice of the desired genotypes overnight, and the detection of a vaginal plug designated as embryonic day 0.5 (E0.5). Nine to 18 days after vaginal plug detection (E9.5 – E18.5), pregnant dams were anesthetised and sacrificed by cervical dislocation. Uterine horns were removed and washed in PBS. Embryos were removed from the uterus and photographs taken with the Leica DFC320 camera, followed by decapitation (as per ethical guidelines).

2.3.1 Embryo fixation for histology purposes

Embryos and placentas to be used for haematoxylin and eosin staining (H&E) and periodic acid-Schiff staining (PAS) were fixed in 4% paraformaldehyde (PFA) overnight at 4 °C, and then washed twice in PBS and stored in 70% ethanol at RT. Genotyping of yolk sacs was undertaken by myself and whole embryos and bisected placentas were sent to either QIMRB histopathology services, or Dr G. Del Monte Nieto, Victor Chang Cardiac Research Institute, for embedding, sectioning and staining.

2.3.2 Cardiac morphology quantification

ImageJ quantification

Histology slides were scanned to create digital images using an Aperio AT Turbo slide scanner (Leica) at 40X magnification. Left and right ventricle field images were acquired in ImageScope (v12.1.0.5029) and saved as individual files. These files were then imported into ImageJ (v1.50c4) bundled with Java v1.8.0_60 for quantification. Ventricle wall thickness was measured towards the apex of the heart, with a minimum of 5 measurements taken from three serial sections per embryo, all hearts were analysed blind to genotype. The mean thickness for each placenta was plotted using GraphPad Prism 6.

2.4 BISULPHITE SEQUENCING

2.4.1 Bisulphite conversion of DNA

Genomic DNA was purified using the AllPrep Mini kit (Qiagen) (adult tissue, fetal liver) or AllPrep Micro kit (Qiagen) (fetal hearts) according to manufacturer's instructions. Fresh tissue (< 30 mg for AllPrep Mini kit or < 5 mg AllPrep Micro kit)

was homogenised in RLT Plus buffer supplemented with β-mercaptoethanol using a needle and syringe. DNA was eluted in 20 µl Milli-Q water, and the eluate re-run through the column a second time to increase final DNA concentration. Sample concentration was measured spectrophotometrically using a NanoDrop 1000 (Thermo Fisher Scientific), and 1 µg of DNA bisulphite converted using the EpiTect Bisulphite conversion kit (Qiagen), as per manufacturer's instructions. Converted DNA was eluted in 20 µl of Milli-Q water, the eluate re-run through the column a second time to increase final DNA concentration and stored at – 20 °C.

2.4.2 PCR amplification of bisulphite treated DNA

PCR was carried out in a Bio-Rad S1000 Thermal Cycler. Each reaction consisted of approximately 20 ng/µl of template gDNA, 0.1 U of Mango*Taq*™ DNA polymerase (Bioline), 200 µM of dNTPs (dATP, dCTP, dGTP, dTTP), 1.5 mM of $MgCl_2$, 500 nM each of forward and reverse primer for each methylated region, and 5X Mango*Taq*™ PCR reaction buffer. The primary PCR reaction for each of the methylated regions was 2 min of initial denaturation at 95 °C, then 35 cycles of denaturation (95 °C, 30 sec), annealing (various temperatures, refer **Appendix I**, 30 sec) and extension (72 °C, 45 sec), with final extension for 5 min at 72 °C. Primary PCR was performed for each region, and PCR product purified using the NucleoSpin Gel and PCR Cleanup Kit according to manufacturer's instructions, followed by secondary PCR performed using the same conditions as described for primary PCR. PCR products were then resolved on a 1.5% TAE (w/v) agarose gel.

2.4.3 Ligation of nested PCR product

All ligations were performed using the pGEM®-T Easy Vector (Promega) according to manufacturer's instructions. Briefly, each ligation reaction contained purified PCR product (purified as in **Section 2.2.3**) and 5 ng of vector at a molar ratio of 3:1. Ligation reactions were performed using the T4 DNA ligase supplied with the kit and undertaken at 4 °C for 16 hrs in a Bio-Rad S1000 thermal cycler, in preparation for transformation and colony screening (below).

2.4.4 Transformation of competent bacterial cells

All transformation steps were carried out using aseptic techniques. 10 µl of ligation reaction was added to 100 µl competent cells and gently mixed, cells were then left on ice for 30 min. Cells were then heat shocked at 42 °C for 45 s and returned to ice for 2 min. LB broth (500 µl) was then added to the cells and incubated at 37 °C and shaken at 220 rpm for 90 min to allow for recovery and expression of the ampicillin resistance gene by the transformants. Cells were then centrifuged at 1,000 rpm for 10 min and 500 µl of supernatant removed. The remaining 100 µl was resuspended and plated onto LB agar/ampicillin (100 µg/ml) plates with IPTG and X-gal solution added just prior to plating out transformed cells (discussed below) and incubated at 37 °C overnight.

2.4.5 Blue/White colony screening

The pGEM®-T Easy Vector (Promega) contains the coding sequence of the *lacZ* gene which can be used to detect the presence of an insert. A vector that has re-ligated with no insert present will express β-galactosidase in the presence of IPTG; β-galactosidase converts X-gal into a blue product, turning colonies with no insert present blue. The presence of an insert disrupts this expression resulting in white

colonies. IPTG (50 µl of 100 mU) and X-gal (20 mg/ml) were plated onto LB agar/ampicillin (100 µg) plates. White colonies were then selected to undergo sequence analysis.

2.4.6 PCR and sequencing of bacterial colonies

PCR was carried out in a Bio-Rad S1000 Thermal Cycler. Each reaction consisted one white colony, 0.1 U of Mango*Taq*™ DNA polymerase (Bioline), 200 µM of dNTPs (dATP, dCTP, dGTP, dTTP), 1.5 mM of $MgCl_2$, 500 nM each of T7 and Sp6 primer, and 5X Mango*Taq*™ PCR reaction buffer. Primer sequences are shown in **Table 2.1**. The PCR reaction was 2 min of initial denaturation at 94 °C, then 35 cycles of denaturation (94 °C, 30 sec), annealing (55 °C, 30 sec) and extension (72 °C, 1 min), with final extension for 5 min at 72 °C. PCR products were resolved on a 1.5% TAE (w/v) agarose gel and products of the correct size then underwent Sanger sequencing as described in **Section 2.2.5**.

2.4.7 Bisulphite sequencing analysis

Analysis of bisulphite sequencing data was undertaken using the BiQ Analyser (Max Planck Institute) (Bock et al., 2005). Sequences were excluded that had a conversion rate of less than 90% and less than 80% homology to the unconverted sequence. For multiple sequences that resembled the same clone, only one sequence was included in analysis. Results from BiQ Analyser program were saved as .html files and methylated and unmethylated CpGs calculated for each biological replicate.

2.4.8 Preparation of competent bacterial cells

All subcloning was performed using competent *E. coli* DH5-α cells. Batches of competent cells were prepared at one time-point and aliquots stored at – 80 °C prior to use. Each of the following steps was undertaken using aseptic techniques and all solutions sterilised prior to use. Cultures of 100 ml of SOB broth were inoculated with 0.5 ml of an overnight culture of *E. Coli* DH5-α and incubated at 37 °C with vigorous aeration until an optical density at 550 nm of 0.3 had been attained. Cultures were rapidly cooled on ice and the remaining steps performed at 4 °C. Bacterial cells were pelleted by centrifugation at 2,000 rpm for 10 min and the supernatant medium decanted. Cells were then gently washed and resuspended in 10 ml TFB I solution and left on ice for 15 min. Cells were then centrifuged as above, and supernatant decanted and the cells gently resuspended in 4 ml TFB II solution. Aliquots of 100 µl of competent cells were stored at – 80 °C.

2.5 FLOW CYTOMETRY ANALYSIS OF TAIL BLOOD

For the collection of blood for flow cytometry, 3 week old mice were anesthetised with isoflurane and 0.2 cm tail biopsied for genotyping. A drop of blood from the tail was also collected in 1 ml of Osmosol (POCD) and mixed by inversion. A Guava® easyCyte™ HT flow cytometer (Millipore) was used to detect and quantify levels of cellular fluorescence. Two channels were used: 488 nm to detect fluorescence from the transgene, and 550 nm used when required to gauge the autofluorescence of cells. The use of two channels allows for a greater accuracy in distinguishing GFP-expressing and non-expressing cell populations (Rasko et al., 1999). Flow cytometry data were analysed using Cell Quest Pro software (Becton

Dickinson) using conservative gates set to exclude 99.9% of wild-type autofluorescing erythrocytes.

2.6 RNA SEQUENCING ANALYSIS

Total RNA was extracted using the AllPrep Mini and Micro kits (Qiagen) as described in **Section 2.2.7**. RNA purity was checked on a Nanodrop 1000 spectrophotometer (Thermo Fisher scientific), and quality tested using the Agilent Bioanalyser RNA 6000 Nano kit (Agilent Technologies).

2.6.1 Library Preparation and Sequencing

Library preparation and Illumina sequencing was performed by the Australian Genome Research Facility (AGRF). Four *Rlf* $^{+/+}$, four *Rlf* $^{MommeD28/MommeD28}$ and three *Rlf* $^{MommeD28/+}$ fetal hearts were sequenced on an Illumina HiSeq 2000 with 50 bp single end reads, and sequenced on two lanes. The Illumina CASAVA1.8 pipeline was used to generate sequencing data.

2.6.2 RNA-seq read mapping and differential expression testing

RNA-seq read mapping for E14.5 liver samples was undertaken by Dr H. Oey. All RNA-seq mapping for E13.5 heart samples was undertaken by myself using the open-web based platform Galaxy (www.galaxy-qld.genome.edu.au) (Blankenberg et al., 2001; Giardine et al., 2005; Goecks et al., 2010). All sequenced reads were aligned to the mouse genome (NCBIM37/mm9) using Tophat (Trapnell et al., 2009), a program that aligns RNA-seq reads to the genome using the short read aligner BowTie, v0.6, and analyses the mapped results to identify splice junctions. The

following Tophat, v0.9, parameters were used -- maximum intron length between donor/acceptor site must be less than 100,000 bp -- coverage search disabled -- read alignments with > 2 mismatches discarded -- library type is unstranded

Aligned reads were then used for differential gene expression testing using default parameters in each of the following programs: CuffDiff, v2.2.1.2 (Trapnell et al., 2012), DESeq, v3.3 (Anders & Huber, 2010) and EdgeR, v2.4.6 (Robinson et al., 2010). Read counts for mRNA transcripts were extracted from the mapped reads using htseq-count (Anders & Huber, 2010) using the parameters --not from a strand specific assay --model for counting reads over the supplied gene model set to intersection strict

Differential expression analysis in EdgeR and DESeq was assessed using the htseq-count matrix as the input matrix and gene annotations from Ensembl (www.ensembl.org, release 67). Default parameters for both programs. Differential expression analysis in CuffDiff was undertaken utilising the reference transcriptome to aggregate read counts per gene, and default parameters used in analysis.

2.6.3 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA) integrates gene expression data with biological information derived from published literature and public databases to predict functional linkages. E14.5 liver differentially expressed genes with an adjusted *p* value >0.05 and minimum fold change of 1.2 were analysed for their direct relationship to a network.

The network score (*p* score) is the probability of finding *n* number of focus genes in a set of a random gene set from the Global Molecular Network, and is calculated with the right-tailed Fishers exact test. The *p* score = -log10 (*p* value).

E13.5 heart genes identified as differentially expressed (> 1.5 fold change, $p \leq 0.05$) were uploaded to IPA and a core biological pathway analysis performed to identify molecular networks and upstream regulators. GOplot was used in order to better visualise the connections between the genes and the canonical pathways highlighted in the Ingenuity pathway and upstream regulator analysis (Walter et al., 2015).

2.7 WHOLE GENOME BISULPHITE SEQUENCING

2.7.1 Sample preparation

Fetal liver was collected from two *Rlf* $^{+/+}$ and two *Rlf*$^{MommeD2/MommeD28\ E14.5}$ embryos derived from the same litter. Tissue was then sent to the Centro National de Analisis Genomico (CNAG, Spain) for bisulphite conversion and sequencing.

2.7.2 Library preparation and sequencing

Genomic DNA samples were spiked with λ DNA and sheared by sonication. Libraries were prepared using the TruSeq Sample Prep Kit (Illumina) and underwent two rounds of sodium bisulphite conversion using the EpiTect Bisulphite Kit (Qiagen). 100 bp paired end sequencing was performed on the Illumina HiSeq 2000. Sequencing reads were trimmed for poor quality at their 3' ends using the program trim_seq (www.biopieces.org) with the options -l 3 -m 3.

2.7.3 Mapping and methylation calling

Reads were mapped to mouse genome (NCBI37/mm9) after first converting known FVB SNPs to the FVB state using the tool FastaAlternateReferenceMaker from the GATK package (McKenna et al., 2010). The SNP positions were obtained from Sanger (rsIDdbSNPv137) and coordinates converted from the NCBI38/mm10 to NCBI37/mm9 using the LiftOver tool from UCSC (Kent et al., 2002). Only nucleotide substitutions were to maintain genome length. The lambda genome sequence and the sequence for the GFP transgene (Preis et al., 2003) were added to the genome sequence as separate chromosomes. Sequencing reads were mapped to the new genome using the program Bismark, version 0.7.12, (Krueger & Andrews, 2011) combined with Bowtie2, version 2-2.1.0 (Langmead & Salzberg, 2012), using the options -N 1 -D 30. The resulting files were filtered for PCR duplicates using the program deduplicate_bismark_alignment_output.pl and CpG methylation values were then extracted using the program bismark_methylation_extractor with the options --ignore_r2 2 --counts --bedGraph --paired-end --no_overlap --comprehensive --merge_non_CpG (Krueger & Andrews, 2011). The CpG methylation calls (one for each strand) were merged to a single methylation call for each CpG using custom scripts.

Differentially methylated regions were identified using the R package bsseq (Hansen et al., 2012). The smoothing was carried out with the options ns=20, h=250, maxGap=100,000,000. Loci with coverage ≥8 in at least one biological replicate for wild-type and mutant were retained and Rlf-DMRs were identified using t-statistic quantile cut-offs of 0.01 and 0.99, requiring >10 CpGs per Rlf-DMR and a mean methylation change >15%.

Methylation values used for hierarchical clustering were the weighted averages of CpG methylation within each genomic coordinate. Only coordinates containing at least 8 CpGs with coverage of at least 6 reads across all samples, were used.

2.8 ENRICHMENT OF HISTONE MARKS

To calculate enrichment of histones at Rlf-DMRs the ENCODE sequencing reads (Shen et al., 2012) for both biological replicates, and for the input, for each tissue and histone modification considered (E14.5 liver, E14.5 heart and E14.5 brain), were obtained from the UCSC Genome Browser (Rosenbloom et al., 2013a). Datasets with longer reads were 3' trimmed such that all datasets contained reads of equal length. The reads were subsequently mapped to the mouse reference genome (NCBI37/mm9) using the program Bowtie2 with the options --trim5 6 --local -L 22 -N 1 retaining only uniquely mapped reads. Reads likely to be PCR duplicates were identified and removed using the program MarkDuplicates Picard (http://picard.sourceforge.net). Enrichment of histone modifications were calculated as described in (Hon et al., 2013b), requiring >2-fold enrichment over input for both replicates.

2.9 GENE ONTOLOGY ANALYSES, DNA SEQUENCE MOTIF SEARCH AND CONSERVATION SCORES

Gene ontology analyses of genes proximal to Rlf-DMRs were carried out using the GREAT tool (McLean et al., 2010). The Homer tool, version 4.6 (Heinz et al., 2010), was used to investigate for enrichment of known and novel transcription factor binding motifs at Rlf-DMRs using the options -size given –cpg, and otherwise default parameters. PhastCons conservation scores (Siepel et al., 2005) for placental

mammals, obtained from the UCSC Genome Browser (Karolchik et al., 2014), were used to get average conservation scores for Rlf-DMRs.

2.10 CHROMATIN IMMUNOPRECIPITATION SEQUENCING ANALYSIS

Fetal liver was collected from eight *Rlf* $^{+/+}$ E14.5 embryos that were then divided into four replicates, consisting of two livers in each replicate. Tissue was then sent to ActiveMotif for chromatin extraction and immunoprecipitation using either the commercial Abcam Rlf antibody (AB115011) or a custom designed Rlf antibody.

2.10.1 Library Preparation and Sequencing

Sequencing of Abcam samples and their input control was undertaken on an Illumina HiSeq 2000 with 50 nt reads, whilst the Ab1 samples and their input (unprecipitated DNA) control were sequenced in the Illumina NextSeq 500 with 75 nt reads.

2.10.2 ChIPseq read mapping and peak calling

All reads were mapped to the mouse genome (NCBI37/mm9) using the Burrows-Wheeler Aligner (BWA) algorithm with default settings (Li & Durbin, 2009). Reads passing Illuminas purity filter, aligned with no more than 2 mismatches and map uniquely to the genome were used in subsequent analyses. Samples from each experiment were normalised alongside of their respective input controls to the same number of unique alignments following the removal of duplicate reads. Peaks were then identified using the MACS v1.4.2 peak finding algorithm (Zhang et al.,

2008). Different *p* value cut offs were used for each experiment (*p* = 1e-7 Ab1, and *p* = 1e-5 Abcam).

2.10.3 ChIPseq enrichment visualisation

For visualisation of putative Rlf binding sites ChIPseq data was loaded onto the UCSC genome browser (www.genome.ucsc.edu.au). MACs generates a peaks.bed file with peak locations that is suitable for uploading into the UCSC genome browser as a custom track. Putative Rlf binding sites were visualised on the *mus musculus* NCBI37/mm9 assembly.

2.11 PAVIS CHIPSEQ ANNOTATION

PAVIS is a tool that enables the annotation of ChIPseq data to genomic regions (http://manticore.niehs.nih.gov/pavis2/). Location of the 200 ChIPseq peaks identified in experiments using two independent Rlf antibodies was input into PAVIS as a .bed file and the parameters: mouse/mm9 refGene set, upstream length 2500 and downstream length 2500 used. Peak enrichment was calculated by PAVIS using binomial testing to compare relative peak enrichment in different genomic regions. With the binomial *p* value equalling the probability of having peaks in a particular genomic region (Huang et al., 2013).

2.12 STATISTICAL ANALYSIS

2.12.1 Standard Student's t-test

The Standard Student's t-test assessed whether the means of two groups are statistically different from each other. A *p*-value of 0.05 implies that there is only a

5% probability that the two groups have statistically different means by chance when they are in fact the same. If the *p*-value is 0.05 or less one can consider that the two groups have been drawn from two populations with statistically significantly different means. Unless otherwise stated, all statistics used in this thesis are two-tailed t-tests with unequal variance, and results are presented as mean ± standard error of the mean (SEM).

2.12.2 F-test

The F-test assesses whether the variances observed in two sets of data are statistically different from each other. A *p*-value of 0.05 implies that there is only a 5% probability that the two groups have statistically different variances by chance when they are in fact the same. If the *p*-value is 0.05 or less one can consider that the two groups have been drawn from two populations with statistically significantly different variances. This is taken into account when calculating the T-test statistic.

2.12.3 Statistical analysis of clonal bisulphite sequencing date

The odds ratio Chi-squared test was used to measure the significance of methylation between genotypes. When conditions for the Chi-squared test were not met (less than 5 replicates) Fisher's exact test was used.

2.12.4 Pearson correlation testing

Relationships between qRT-PCR and RNA-seq, and litter size and pup weight, were determined by Pearson correlation analysis using GraphPad Prism6. R^2 values were considered significant when a two tailed *p* value was ≤ 0.05.

Chapter 3: Analysing the effect of loss of RLF in E14.5 fetal liver

3.1 INTRODUCTION

Previous work by the Whitelaw laboratory has shown that loss of Rlf results in a decrease in the number of erythrocytes expressing the GFP reporter transgene, an increase in methylation at the transgene, and haploinsufficiency for *Rlf* increased the probability of silencing at the A^{vy} allele (Daxinger et al., 2013). To gain insight into how loss of *Rlf* impacts upon transcription and methylation across the whole genome RNA-seq and whole genome bisulphite sequencing were undertaken[1]. The identification of direct and indirect endogenous targets of Rlf may suggest cellular pathways and functions that are regulated by Rlf. The identification of differentially methylated regions in the genome of *Rlf* homozygous mutants will identify regions in the genome that require Rlf for maintenance of methylation. Additionally, some but not all, of the zinc fingers present in Rlf have been found to be capable of directly binding DNA (Harten et al., 2015). ChIPseq was undertaken to identify putative Rlf binding sites in the genome using two independent Rlf antibodies. Fetal liver was chosen for this study as at E14.5 this is the hematopoietic organ for the embryo and the ENU mutagenesis screen, from which Rlf was identified, assesses transgene variegation in red blood cells. Encyclopedia of DNA elements (ENCODE) data of many functional elements is also publicly available for comparison of these data sets to our own.

[1] A significant portion of results from this chapter have been published in Harten et al., 2015.

3.2 RESULTS

3.2.1 Loss of Rlf has no impact on fetal liver morphology

Previous experiments have shown Rlf to be expressed in the fetal liver, although no further experiments investigating the impact of loss of Rlf on the liver have been undertaken (Harten et al., 2015). To determine whether loss of Rlf alters cell morphology in fetal livers I undertook histological analysis from fixed whole *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ E14.5 embryos (n = 5 per genotype). No overt differences in cell morphology were observed when comparing *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal livers following haematoxylin and eosin (H&E) staining, **Figure 3.1A**. Western blot analysis of Rlf using the custom designed antibody (refer **Section 4.2.1**), Ab1, in *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ E14.5 fetal livers also showed Rlf is expressed in wild-type but not homozygous livers, **Figure 3.1B**. This confirms the liver as a suitable tissue to investigate the effect of loss of Rlf on gene expression and methylation.

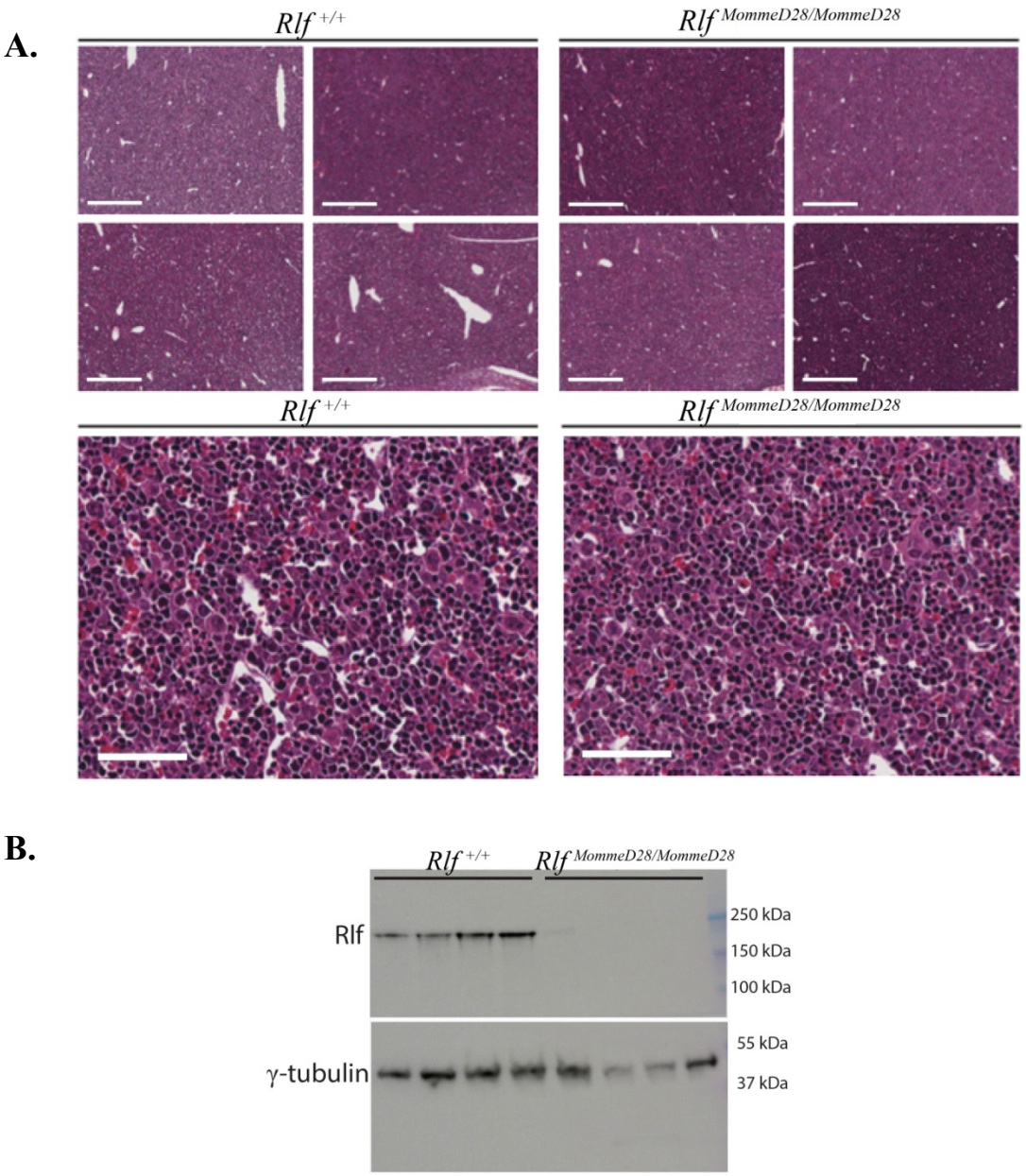


Figure 3.1: Rlf is expressed in the fetal liver and has no impact on morphology.
A. Haematoxylin and eosin staining of E14.5 *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal liver. Whole embryos were fixed in 4% PFA and embedded in paraffin. Sections were cut and stained by QIMR Berghofer Histotechnology facility, representative sections are shown. n = 5 per genotype. Scale bars, 300 μm (top) and 60 μm (bottom) respectively. **B.** Western blotting for Rlf in protein extracted from E14.5 *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal livers. Each lane represents an individual embryo and littermates were used for analysis. Rlf was detected at 250 kDa and γ-tubulin used as a loading control.

3.2.2 Rlf regulates endogenous gene expression in the fetal liver

Previous studies by the Whitelaw laboratory have shown loss of *Rlf* alters expression of the GFP transgene in mutant mice (Daxinger et al., 2013). The effect of loss of *Rlf* on endogenous gene expression has not previously been undertaken. Thus I chose to undertake RNA sequencing (RNA-seq) to investigate the effect of loss of *Rlf* on the fetal liver transcriptome. The identification of genes whose expression is altered in *Rlf* mutants may suggest cellular pathways and functions that are regulated by Rlf, which in turn may point towards how Rlf acts as an epigenetic modifier. RNA-seq analysis of E14.5 liver extracted from $Rlf^{+/+}$, $Rlf^{MommeD28/+}$ and $Rlf^{MommeD28/MommeD28}$ littermates (n = 3 per genotype). Briefly, mRNA-seq library preparation and Illumina HiSeq 2000 sequencing with 50 bp single end reads was performed by Australian Genome Research Facility (AGRF). Each fetal liver underwent sequencing and alignment to the mouse genome (NCBI37/mm9) individually. Read counts were then extracted from the mapped reads for each sample using HTSeq-count and a matrix file of read counts for all samples generated for differential expression testing (Anders et al., 2015). Analysis of differentially expressed genes was then undertaken in DEseq, using default parameters and all p values are adjusted for multiple testing using the Benjamini Hochberg method (Anders & Huber, 2010; Benjamini & Hochberg, 1995).

Using an adjusted p value cut off of $p < 0.05$ and fold change of > 1.2, 184 genes were found to be differentially expressed in $Rlf^{MommeD8/MommeD28}$ compared to $Rlf^{+/+}$ E14.5 liver (97 up-regulated and 87 down-regulated). The top 20 up- and down-regulated genes in $Rlf^{MommeD28/MommeD28}$ fetal liver, based on fold change, are shown in **Table 3.1** and **Table 3.2**, and **Appendix II** lists all of the RNA-seq data

comparing *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ genotypes. Analysis of genes whose expression was altered greater than two fold in *Rlf* $^{MommeD28/MommeD28}$ fetal liver found approximately four times as many genes were down-regulated (n = 31) than up-regulated (n = 7),

Table 3.3**.** This correlates with previous finding of *Rlf* being required for gene activation (Daxinger et al., 2013). *Rlf* was not found to be differentially expressed following RNA-seq analysis. Further investigation of RNA-seq mapped data found reads mapped to intron 2 as well as no reads mapping to the beginning of exon 3 potentially due to alternate splicing. A representative *Rlf* $^{MommeD28/MommeD28}$ sample illustrating this mapping is provided in **Figure 3.2.** That we see no Rlf protein expression in the fetal liver (**Figure 3.1**) suggests the RNA product may undergo degradation, for example via nonsense mediated decay.

Comparison of *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/+}$ RNA-seq data found no genes to be significantly differentially expressed between the two genotypes following correction for multiple testing, **Appendix III**. Although no genes were identified as being statistically differentially expressed, investigation of gene expression in *Rlf* $^{MommeD28/+}$ and *Rlf* $^{MommeD28/MommeD28}$ livers compared to wild-type revealed read counts for genes in *Rlf* $^{MommeD28/+}$ were altered in the same direction as those observed in *Rlf* $^{MommeD28/MommeD28}$, **Table 3.4**.

Table 3.1: Top 20 down-regulated genes in *Rlf*$^{MommeD28/MommeD28}$ fetal liver.

Symbol	Gene Name	Fold Change	Adjusted P value
Hpd	4-hydroxyphenylpyruvate dioxygenase	0.011	1.90E-19
Cyp2d37-ps	Cytochrome P450, family 2, subfamily d, polypeptide 37, pseudogene	0.024	5.68E-05
Tcfl5	Transcription factor-like 5 (basic helix-loop-helix)	0.097	4.32E-02
Hormad2	HORMA domain containing 2	0.100	5.96E-03
Mc2r	Melanocortin 2 receptor	0.193	7.05E-04
Prss50	Protease, Serine, 50	0.211	4.16E-20
Atp2b2	ATPase, Ca++ transporting, plasma membrane 2	0.263	8.82E-04
Upb1	Ureidopropionase, beta	0.272	1.14E-20
Pnliprp2	Pancreatic lipase-related protein 2	0.273	3.21E-02
Timd2	T-cell immunoglobulin and mucin domain containing 2	0.279	4.95E-33
Apoc1	Apolipoprotein C-I	0.281	1.96E-14
Psma8	Proteasome subunit alpha type 8	0.283	5.68E-05
Cyp2d9	Cytochrome P450, family 2, subfamily d, polypeptide 9	0.295	5.13E-02
Zfp783	Zinc finger protein 783	0.313	1.42E-05
Dnajb3	DnaJ (Hsp40) homolog, subfamily B, member 3	0.316	8.78E-06
Crispld2	Cysteine-rich secretory protein LCCL domain containing 2	0.332	3.72E-03
Dnahc8	Dynein, axonemal, heavy chain 8	0.350	2.08E-02
Sycp1	Synaptonemal complex protein 1	0.354	5.13E-03
Haao	3-hydroxyanthranilate 3,4-dioxygenase	0.355	7.01E-08
Liph	Lipase, member H	0.359	1.84E-02

Table 3.1: Top 20 down-regulated genes in *Rlf*$^{MommeD28/MommeD28}$ fetal liver compared to wild-type, ordered by fold change, following RNA-seq analysis. Fetal liver was collected from three individual E14.5 embryos for each genotype; all embryos were from the same litter. RNA-seq was performed by AGRF on the Illumina HiSeq 2000 and bioinformatic analysis in DESeq provided by Dr H. Oey. Adjusted *p* value controls for false discovery rate through the use of the Benjamini-Hochberg procedure. (Full data Appendix II).

Table 3.2: Top 20 up-regulated genes in *Rlf*$^{MommeD28/MommeD28}$ fetal liver.

Symbol	Gene Name	Fold Change	Adjusted P value
Mgam	Maltase-glucoamylase	3.497	3.80E-18
Myo5c	Myosin VC	3.224	2.15E-03
Nupr1	Nuclear protein transcription regulator 1	2.827	3.60E-02
Sox9	SRY (sex determining region Y)-box 9	2.597	3.02E-04
Nefh	Neurofilament, heavy polypeptide	2.310	1.84E-02
Clic6	Chloride intracellular channel 6	2.20	9.80E-03
Aldh1a7	Aldehyde dehydrogenase family 1, subfamily A7	1.995	2.04E-05
Igsf11	Immunoglobulin superfamily, member 11	1.984	5.68E-05
Myof	Myoferlin	1.931	3.39E-02
Vldlr	Very low density lipoprotein receptor	1.788	5.68E-03
Akr1c12	Aldo-keto reductase family 1, member C12	1.782	2.11E-05
Pcdh7	Protocadherin 7	1.781	6.79E-03
Zbtb20	Zinc finger and BTB domain containing 20	1.741	2.15E-02
Eml5	Echinoderm microtubule associated protein like 5	1.720	2.19E-02
Dak	Dihydroxyacetone kinase 2 homolog (yeast)	1.683	4.56E-09
Card6	Caspase recruitment domain family, member 6	1.652	3.72E-03
Ppm1e	Protein phosphatase 1E (PP2C domain containing)	1.644	6.92E-04
Tnfsf14	Tumor necrosis factor (ligand) superfamily, member 14	1.642	3.14E-02
Kcnj5	Potassium inwardly-rectifying channel, subfamily J, member 5	1.639	6.54E-06
Scai	Suppressor of cancer cell invasion	1.611	5.68E-05

Table 3.2: Top 20 up-regulated genes in *Rlf*$^{MommeD28/MommeD28}$ fetal liver compared to wild-type, ordered by fold change, following RNA-seq analysis. Fetal liver was collected from three individual E14.5 embryos for each genotype; all embryos were from the same litter. RNA-seq was performed by AGRF on the Illumina HiSeq 2000 and bioinformatic analysis in DESeq provided by Dr H. Oey. Adjusted *p* value controls for false discovery rate through the use of the Benjamini-Hochberg procedure. (Full data Appendix II).

Table 3.3: More genes are down-regulated in *Rlf*$^{MommeD28/MommeD28}$ fetal liver than up-regulated.

Fold Change	Differentially regulated genes in *Rlf*$^{MommeD28/MommeD28}$ fetal liver (n)	
≥ 4	↑ 0	↓ 2
≥ 3	↑ 2	↓ 10
≥ 2	↑ 7	↓ 31
≥ 1.5	↑ 31	↓ 59

Table 3.3: The number of genes significantly ($p < 0.05$) differentially expressed in *Rlf*$^{MommeD28/MommeD28}$ E14.5 livers relative to *Rlf*$^{+/+}$ controls at various fold-change cut-offs, shows more are down-regulated than up-regulated. ↑ and ↓ indicates increased or reduced expression in *Rlf* mutants.

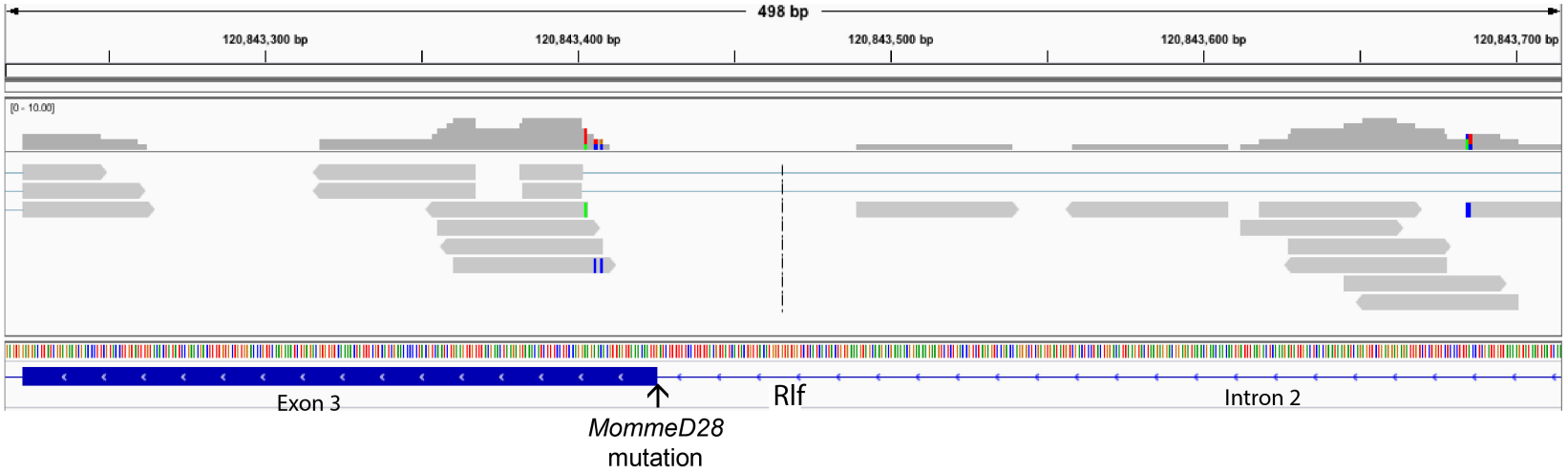


Figure 3.2: RNA-seq mapped reads in a representative *Rlf*$^{MommeD28/MommeD28}$ sample.

Representative RNA-seq mapping of a *Rlf*$^{MommeD28/MommeD28}$ replicate, showing a large number of reads mapped to intron 2 of *Rlf* and no reads mapped to the beginning of Exon 3 where the *MommeD28* mutation is located.

Table 3.4: Gene expression in *Rlf*$^{MommeD28/+}$ fetal liver is in the same direction as *Rlf*$^{MommeD28/MommeD28}$ fetal liver.

Symbol	Gene Name	↑ or ↓in *MommeD28/MommeD28*	↑ or ↓in *MommeD28/+*	*P* value in *MommeD28/+*	Adjusted *P* value in *MommeD28/+*
Timd2	T-cell immunoglobulin and mucin domain containing 2	↓	↓	3.47E-05	0.270011
Itga6	Integrin alpha 6, full insert sequence	↑	↑	3.53E-05	0.270011
Pgp	Phosphoglycolate phosphatase	↓	↓	3.91E-05	0.270011
Ipcef1	Interactor protein for cytohesin exchange	↓	↓	4.92E-05	0.270011
Psma8	Proteasome subunit alpha type-7-like	↓	↓	1.56E-04	0.554285
Itga2b	Integrin alpha-IIb	↑	↓	1.67E-04	0.554285
BC100530	Stefin A-like protein	↓	↓	1.95E-04	0.554285
Tubb1	Tubulin, beta 1	↑	↑	2.02E-04	0.554285
Nr2f6	Nuclear receptor subfamily 2, group F, member 6	↓	↓	2.69E-04	0.655303
Dak	Bifunctional ATP-dependent dihydroxyacetone	↑	↑	3.11E-04	0.662043

Table 3.4: Gene expression in *Rlf*$^{MommeD28/+}$ samples is in the same direction as *Rlf*$^{MommeD28/MommeD28}$ gene expression. Significance of changes in expression in *Rlf*$^{MommeD28/+}$ samples before (*P* value*)* and after (adjusted *P* value) correction for multiple testing. (Full data Appendix III).

Given this finding I chose to use RNA extracted from wild-type, heterozygous and homozygous *MommeD34* mice, an independent line, for validation. A selection of genes whose expression had a fold change greater than three and were significantly ($p < 0.005$) up- or down-regulated in $Rlf^{MommeD28/MommeD28}$ fetal liver were chosen for quantitative real-time RT-PCR validation (qRT-PCR), **Figure 3.A-B**. mRNA expression was normalised to the housekeeping gene β-actin as expression values were found to not vary between genotypes. qRT-PCR validation found each gene to also be significantly differentially expressed in $Rlf^{MommeD34/MommeD34}$ fetal liver. These results indicate that the differential expression of these genes is due to loss of Rlf, although which genes are direct versus indirect targets is still unclear. Significant differences in mRNA expression comparing $Rlf^{MommeD34/+}$ and $Rlf^{+/+}$ fetal livers were also observed, unlike in the RNA-seq data. A number of factors may contribute to the differences observed between the RNA-seq and qRT-PCR methods. These include bias introduced during library preparation, variability between replicates, or correction for multiple testing.

To investigate this further I examined the *p* value and adjusted *p* value from RNA-seq data comparing $Rlf^{+/+}$ and $Rlf^{MommeD28/+}$ E14.5 liver. While no genes were found to be significant following correction testing (adjusted *p* value), 414 genes were found to be significantly differentially expressed according to their non-adjusted *p* value ($p < 0.05$), the top 10 of which are shown in **Table 3.4** (**Appendix III)**. Analysis of the fold change values from the RNA-seq and qRT-PCR datasets showed they correlated well with each other ($R = 0.8133$), **Figure 3.3**, suggesting the correction for multiple testing is the reason for discrepancy between the two

methods. Confirmation of our findings in tissue from an independent *Rlf* mutant line provides strong evidence that the validated genes are regulated by *Rlf*.

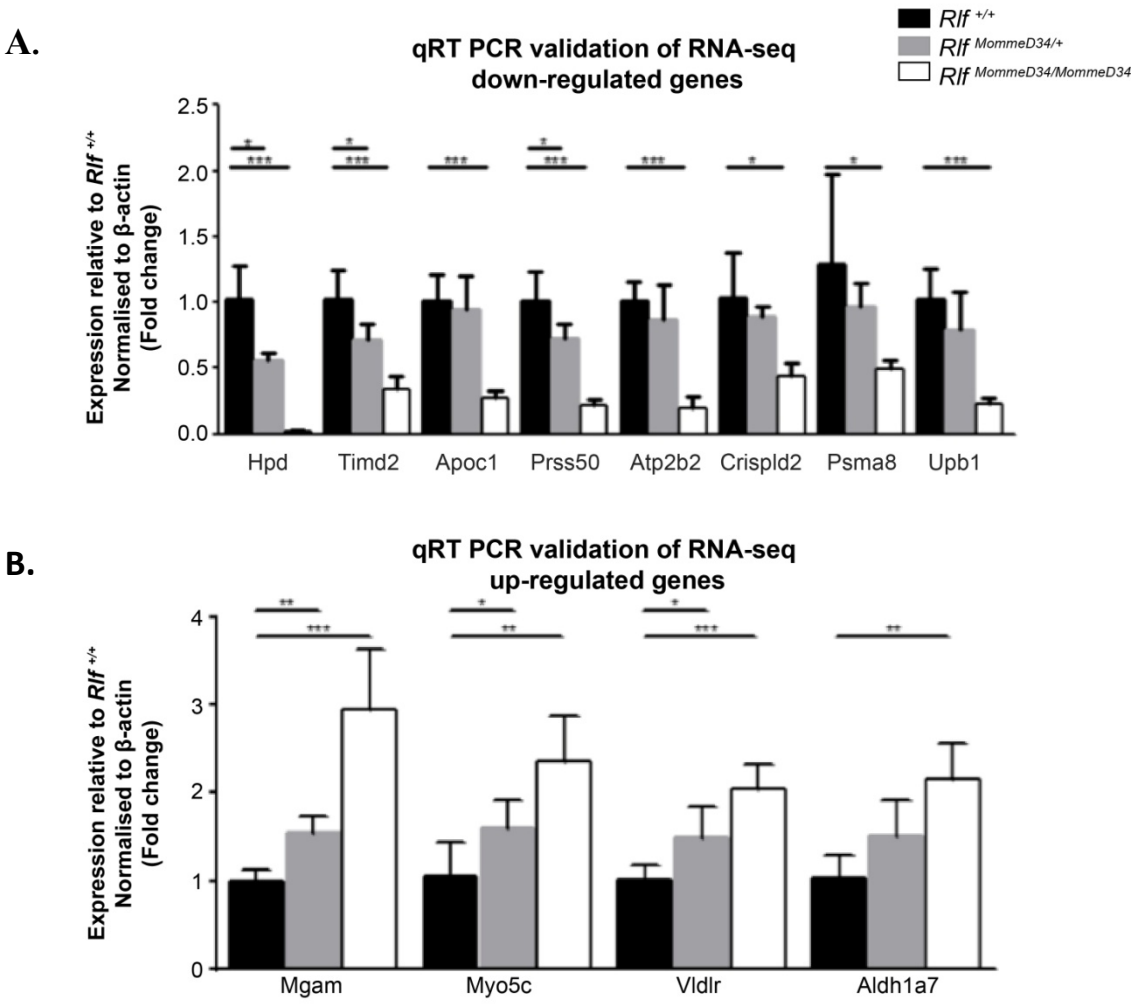


Figure 3.3: Validation of differentially expressed genes from RNA-seq.

A, B. qRT-PCR analysis of differentially expressed genes identified from RNA-seq was performed on RNA extracted from the independent *Momme34* mouse line. mRNA expression of the most significantly up and down-regulated genes according to RNA-seq results was investigated. mRNA expression was normalised to β-actin and wild-type controls. Error bars represent SEM and Student's t-test was used to calculate *p* values (**p* < 0.05, ***p* < 0.005, ****p* < 0.0005). *n* = min. 4 for each genotype.

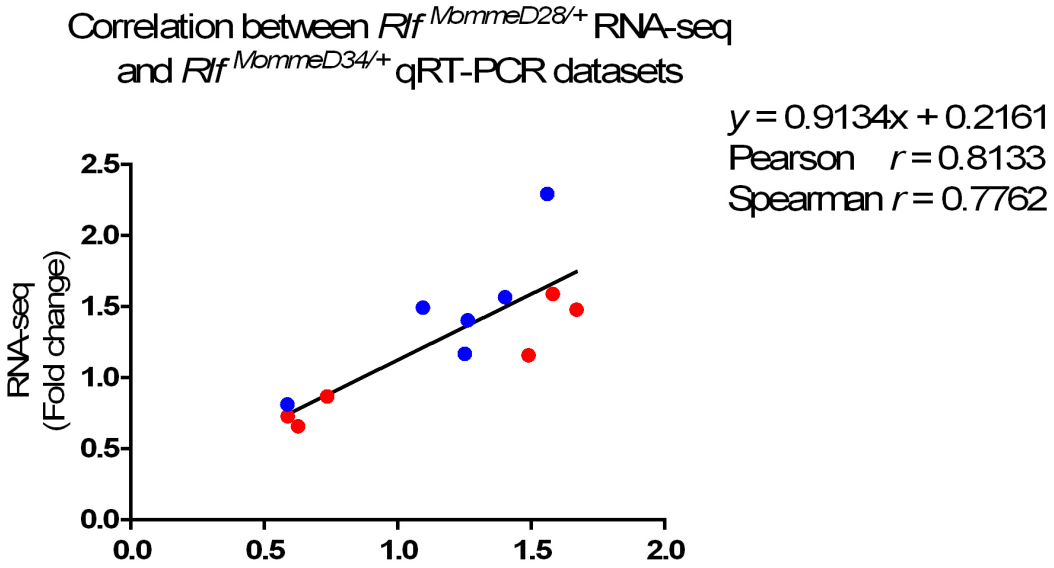


Figure 3.3: Correlation of *Rlf*$^{MommeD28/+}$ RNA-seq and *Rlf*$^{MommeD34/+}$ qRT-PCR datasets

Rlf $^{MommeD28/+}$ RNA-seq (y-axis) and *Rlf* $^{MommeD34/+}$ qRT-PCR (x-axis) datasets correlated well with each other, using the fold change measure of genes chosen for qRT-PCR validation. Genes included in the analysis include those found to be significantly differentially expressed ($p < 0.05$) following qRT-PCR but not RNA-seq (red), and those not significantly differentially expressed following either method (blue).

Next I investigated whether genes identified as being significantly differentially expressed in *Rlf*$^{MommeD28/MommeD28}$ fetal liver were associated with the same biological pathways and processes. A small number of genes were found to have a minimum fold change of 1.5, and an adjusted *p* value of < 0.05 (90 genes in total). To increase the ability to identify biological mechanisms and pathways that may be altered in *Rlf* mutants, this gene list was expanded to include genes with a minimum fold change of 1.2, and an adjusted *p* value of < 0.05 (184 genes in total).

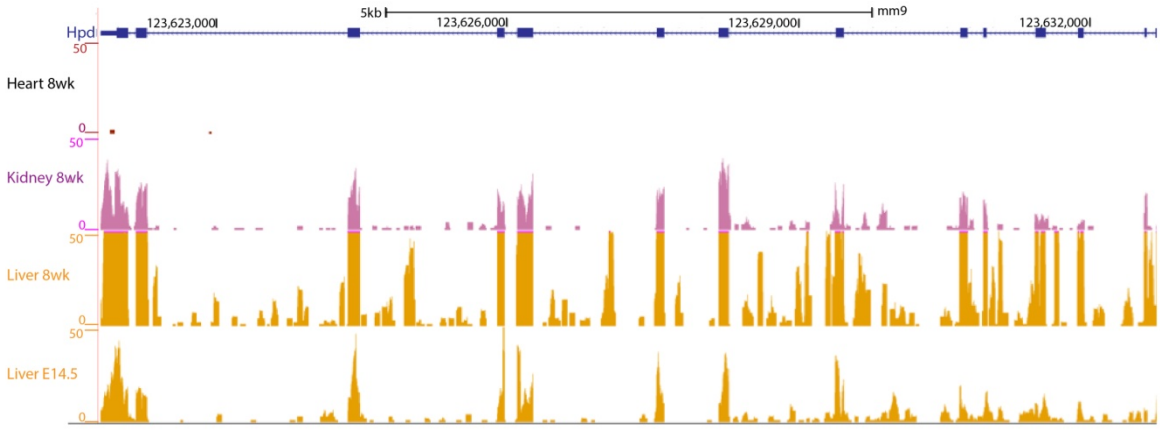
Analysis of these 184 differentially expressed genes using QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity), revealed the top associated network functions to be lipid metabolism (*p* score = 19), e.g. *Aldh1a7* (Aldehyde dehydrogenase family 1, subfamily A7), and cardiovascular system development and function (*p* score = 52), e.g. *Crispld2* (Cysteine-rich secretory protein LCCL domain containing 2). This analysis is consistent with the role of the liver in metabolism and fetal haematopoiesis (Lee et al., 2012).

3.2.3 Investigating Hpd and the Fetal Liver

Of the genes found to be differentially expressed in *Rlf* mutant fetal liver I chose to investigate *Hpd*, the most highly down-regulated gene, in more detail. Expression of *Hpd* mRNA was found to be decreased almost 94 fold in homozygous fetal liver versus wild-type following RNA-seq analysis and validated using qRT-PCR in an independent *Momme* line (**Table 3.1** and **Figure 3.**).

In both mice and humans Hpd protein is primarily expressed in the liver, but is also detected in the kidneys and is required for the catalysis of 4-hydroxyphenylpyruvate to homogentisate in the tyrosine catabolism pathway, which is conserved between humans and mice, **Figure 3.4A-B** (Kent et al., 2002; Tanaka et al., 2006). Loss of Hpd function results in a build up of tyrosine (Endo et al., 1991). Mutations in *HPD* in humans have been shown to lead to two disorders – Tyrosinemia Type III (autosome recessive/ homozygous mutation) and Hawkinsinuria (dominant/ heterozygous mutation) (Tomoeda et al., 2000). In both disorders patients present with mental retardation and show "a failure to thrive" (Borden et al., 1992; Cerone et al., 1997; Ellaway et al., 2001; Niederwieser et al., 1977). As initial observations of *Rlf* mutants found they weighed less than wild-type littermates in late gestation (refer **Section 4.2.6**) (Daxinger et al., 2013), this raised the question whether *Hpd* deficiency in *Rlf* mutant mice contributes to the reduced weight phenotype observed.

A.



B. **Tyrosine Catabolism Pathway**

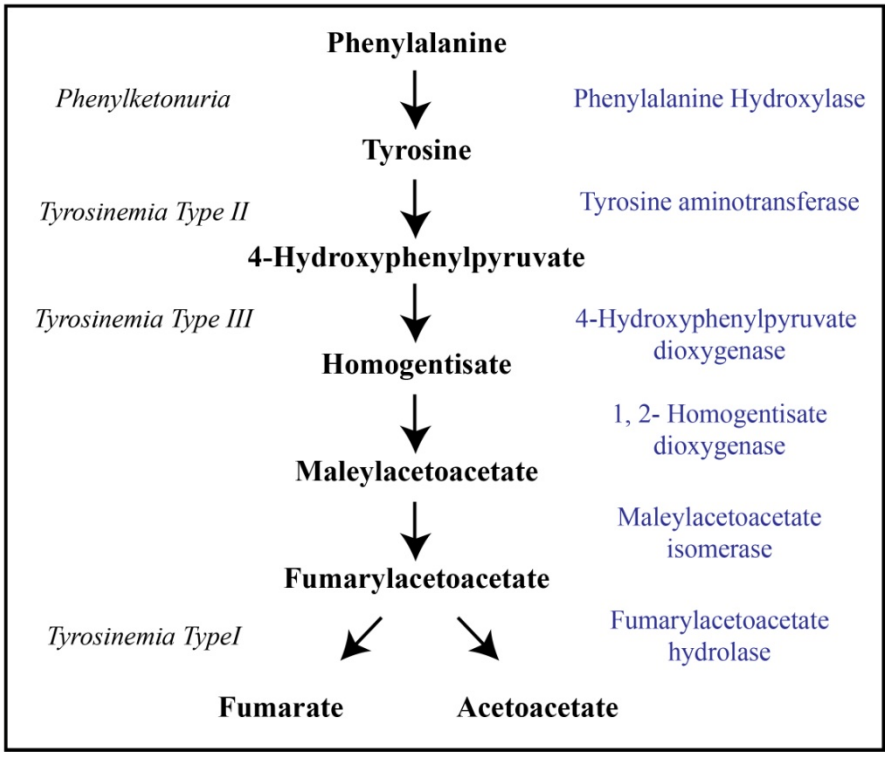


Figure 3.4: *Hpd* expression in the liver and tyrosine catabolism pathway.

A. UCSC Genome Browser snapshot (http://genome.ucsc.edu.au) (NCBI37/mm9), showing *Hpd* (blue) RNA-seq ENCODE expression data in different tissues – heart (red), kidney (pink), and liver (yellow), at different time-points – adult (8 weeks) and embryo (E14.5). **B.** Tyrosine catabolism pathway, enzymes that catabolise each step are shown on the right and diseases occurring due to deficiency in an enzyme shown on the left. Hpd identified as being down-regulated in *Rlf* mutant livers is highlighted. Adapted from Tanaka et. al., 2006.
(Appendix XIV).

Mice lacking Hpd have been found to have increased tyrosine levels in the blood (Endo et al., 1991). As RNA-seq and qRT-PCR validation in *MommeD28* and *MommeD34 Rlf* mutants showed a significant decrease in *Hpd* expression (94 fold (**Figure 3.**), and qRT-PCR of *Hpd* in E14.5 *Rlf* $^{MommeD8/MommeD34}$ offspring also showed a reduction in *Hpd* mRNA in mice with reduced *Rlf* expression (results presented and discussed in **Chapter 5,**Figure 5.3). I asked whether loss of *Rlf* affects tyrosine levels in the blood of *Rlf* mutant mice. To study this I utilised hypomorphic compound heterozygous adult mice (*Rlf* $^{MommeD8/MommeD34}$) and age matched wild-type mice (n = 5 per genotype), and also E18.5 *Rlf* $^{MommeD34/MommeD34}$ and *Rlf* $^{+/+}$ littermates (n = minimum 3 per genotype). Compound heterozygous mice were chosen for the adult portion of this study as homozygous *MommeD28* and *MommeD34* mice die shortly after birth, making it unfeasible to perform the study in adult mice from these lines (refer **Sections 4.2.5** and **4.2.6**) (Daxinger et al., 2013). Compound heterozygous mice have also been found to display a reduced weight phenotype similar to the null *Rlf MommeD28* and *MommeD34* lines (refer **Figure 5.4**). I collected whole blood from adult mice via cardiac puncture following euthanasia with CO_2, and from embryos following decapitation. Blood was allowed to clot at room temperature and centrifuged at 10,000 rpm for 5 minutes; serum was removed and sent for tyrosine analysis to Dr. B. McWhinney, Queensland Health.

No significant difference (p = 0.325) in tyrosine levels in the blood was observed in adult compound heterozygote mice compared to wild-type mice, **Figure 3.5.** Comparison of the raw data collected from both *Rlf* $^{+/+}$ and *Rlf* $^{MommeD8/MommeD34}$ adult mice (63 -120 µmol/L, Table 3.5) to previously published data by other groups found our results to be similar to that of wild-type mice in other studies (50 – 120

μmol/L), whilst *Hpd*$^{-/-}$ mice had an approximate 10 fold increase in tyrosine in the blood (Endo et al., 1991). This suggests that even though there is reduced mRNA expression of *Hpd* in these mice a minimal amount of Hpd protein may enable the tyrosine pathway to function effectively**.** Western blotting for Hpd using the Aviva Systems Biology antibody (OAAB05183), was unable to detect any signal in either wild-type or mutant liver (data not shown), making it difficult to determine how Hpd protein levels are affected in these mice.

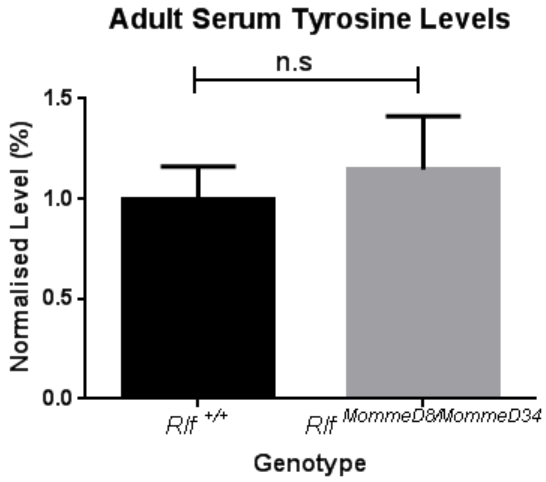


Figure 3.5: Analysis of serum tyrosine levels in adult *Momme* mice.

Analysis of tyrosine levels in serum in adult *Rlf* $^{MommeD8/MommeD34}$ compound heterozygous and wild-type age matched mice. Blood was collected from adults via cardiac puncture and left to clot before centrifuging and removing serum. Tyrosine amino acid assay was performed by Dr B. McWhinney at Qld Health. Tyrosine levels were normalised to the average tyrosine level of wild-type littermates. Error bars represent SEM and Student's t-test was used to calculate *p* values. n = 5 per genotype.

Table 3.5: Adult Mouse Tyrosine values.

Genotype	Tyrosine Level (µmol/L)
Rlf $^{+/+}$	71
Rlf $^{+/+}$	79
Rlf $^{+/+}$	96
Rlf $^{+/+}$	73
Rlf $^{+/+}$	63
Rlf $^{MommeD8/MommeD34}$	87
Rlf $^{MommeD8/MommeD34}$	120
Rlf $^{MommeD8/MommeD34}$	64
Rlf $^{MommeD8/MommeD34}$	87
Rlf $^{MommeD8/MommeD34}$	80

Table 3.5: Raw tyrosine values (µmol/L) of adult mice shows no difference in tyrosine levels between genotypes.

At E18.5 a small but significant (p = 0.0087) decrease in tyrosine in *Rlf* $^{MommeD34/MommeD34}$ embryos compared to *Rlf* $^{+/+}$ was observed, **Figure 3.6** and **Table 3.6**. This is in contrast to the anticipated increase in tyrosine in mice lacking Hpd expression. During this PhD I have found *Rlf* $^{MommeD34/MommeD34}$ embryos weigh approximately 30% less at E18.5 than their wild-type littermates (refer **Section 4.2.6**. This difference in size between the two genotypes could be the cause of the reduced tyrosine levels observed, with smaller embryos either not requiring or producing tyrosine at the same concentration as larger *Rlf* $^{+/+}$ embryos, this will be further discussed in **Section 3.3**. Due to the observation that this difference in size between genotypes occurs early in development (refer **Figure 4.9**), it would be unfeasible to examine tyrosine levels at earlier time-points in embryos which have a reduced blood supply compared to the E18.5 embryos examined here.

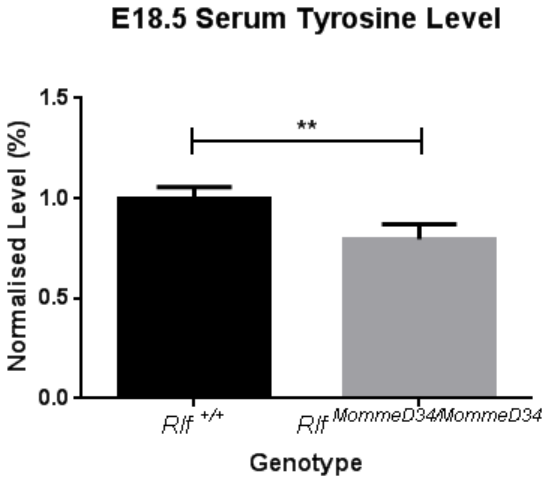


Figure 3.6: Analysis of serum tyrosine levels in embryo *Momme* mice.
Analysis of tyrosine levels in serum in E18.5 *Rlf* $^{+/+}$ and *Rlf* $^{MommeD34/MommeD34}$ embryos. Blood was collected from embryos via decapitation and left to clot before centrifuging and removing serum. Tyrosine amino acid assay was performed by Dr B. McWhinney at Qld Health. Tyrosine levels were normalised to the average tyrosine level of wild-type littermates. Error bars represent SEM and Student's t-test was used to calculate *p* values (**$p < 0.01$). *n* = minimum of 3 per genotype.

Table 3.6: Mouse Embryo Tyrosine value.

Genotype	Tyrosine Level (µmol/L)
Rlf $^{+/+}$	180
Rlf $^{+/+}$	171
Rlf $^{+/+}$	161
Rlf $^{+/+}$	160
Rlf $^{MommeD34/MommeD34}$	125
Rlf $^{MommeD34/MommeD34}$	148
Rlf $^{MommeD34/MommeD34}$	128

Table 3.6: Raw tyrosine values (µmol/L) of embryos shows a slight reduction in tyrosine levels in *MommeD34* homozygous mice.

3.2.4 Analysis of the Effect of Loss of Rlf on the Fetal Liver Methylome

Following the observation of changes in transcription in our mouse modes, I next asked how loss of Rlf impacts methylation in these mice. Loss of Rlf has been shown to increase methylation at the GFP transgene in each of the three *Rlf* mutant mouse lines (Daxinger et al., 2013), however, the effect of Rlf on methylation across the genome is unknown. As a result I decided to utilise genome-wide bisulphite sequencing (GWBS) to investigate methylation across the genome in the *MommeD28* mouse line. GWBS allows for genome-wide analysis of methylation patterns, whilst other methods such as methylated DNA immunoprecipitation and reduced representation bisulphite sequencing are sufficient for analysis of methylation at CpG islands and promoter regions (Meissner et al., 2005; Tost & Gut, 2007; Weber et al., 2005). As it is unknown where, or at which regions, in the genome loss of *Rlf* may alter methylation we chose to use a genome-wide analysis method. GWBS is a more expensive method, however its reproducibility and accurate results allow for a smaller number of replicates to be studied with the same outcome as using more replicates and an alternate methylation analysis method (Tost & Gut, 2007).

Of the nine samples sent for RNA-seq, two $Rlf^{MommeD28/MommeD28}$ and two $Rlf^{+/+}$ littermates from this group were also sent for GWBS. Sequencing was completed for this sample set by Centre Nacional d'Analisi Genomica (CNAG), Spain. Initially, this dataset was used to determine how loss of Rlf impacts methylation across the genome.

Key questions I was interested in addressing from this data were:

1. Does loss of Rlf alter methylation at specific regions across the genome?

2. Do sites identified as Rlf differentially methylated regions (Rlf-DMRs) correlate with genes identified as differentially expressed in the fetal liver?

3. Are the Rlf-DMRs present in *Rlf* $^{MommeD28/MommeD28}$ mice also present in *Rlf* $^{MommeD34/MommeD34}$ mice?

Subsequently bioinformatic analysis to address these questions was carried out by Professor I. Gut, CNAG, and Dr H. Oey, La Trobe University. These findings were reported in Harten et al., 2015, alongside of complementary studies outside of the scope of my thesis, investigating the effect of Rlf on the methylome at different time-points.

Whilst GWBS analysis pipelines were being established by Dr H. Oey, I firstly asked whether I could identify differential methylation at putative regulatory regions that overlapped with genes identified as being differentially expressed in our RNA-seq analysis, and secondly whether this could be validated using bisulphite PCR. DNA from the independent mouse line, *MommeD34*, was bisulphite converted and Sanger sequenced according to standard protocols (refer **Section 2.4**). Two candidate genes that showed significant differences in mRNA expression, *Hpd* and *Prss50* (Protease, Serine, 50), down-regulated ~93 and ~5 fold, respectively (**Table 3.1**), were chosen to determine whether differential expression correlated with differential methylation at putative regulatory regions.

A DNase I hypersensitive site (DHS) upstream of *Hpd* reflecting an enhancer region was identified by manually viewing GWBS data alongside of E14.5 liver DHS produced by the ENCODE project (Rosenbloom et al., 2013b). No other potential differentially methylated regions were found to be proximal to *Hpd*. This region consisted of 5 CpG sites and appeared differentially methylated from GWBS data in $Rlf^{MommeD28/MommeD28}$ (~20%) fetal liver compared to $Rlf^{+/+}$ (~10%), and was subsequently chosen for validation, **Figure 3.7A**.

Many recent publications investigating differential methylation classify a ten percent difference in methylation as being a biologically significant change with reduced probability of representing an artificial methylation change (Stefansson et al., 2015; Yuen et al., 2010; Zeller et al., 2012). In accordance with the literature I have classified candidate regions with a minimum methylation difference of ten percent as being differentially methylated.

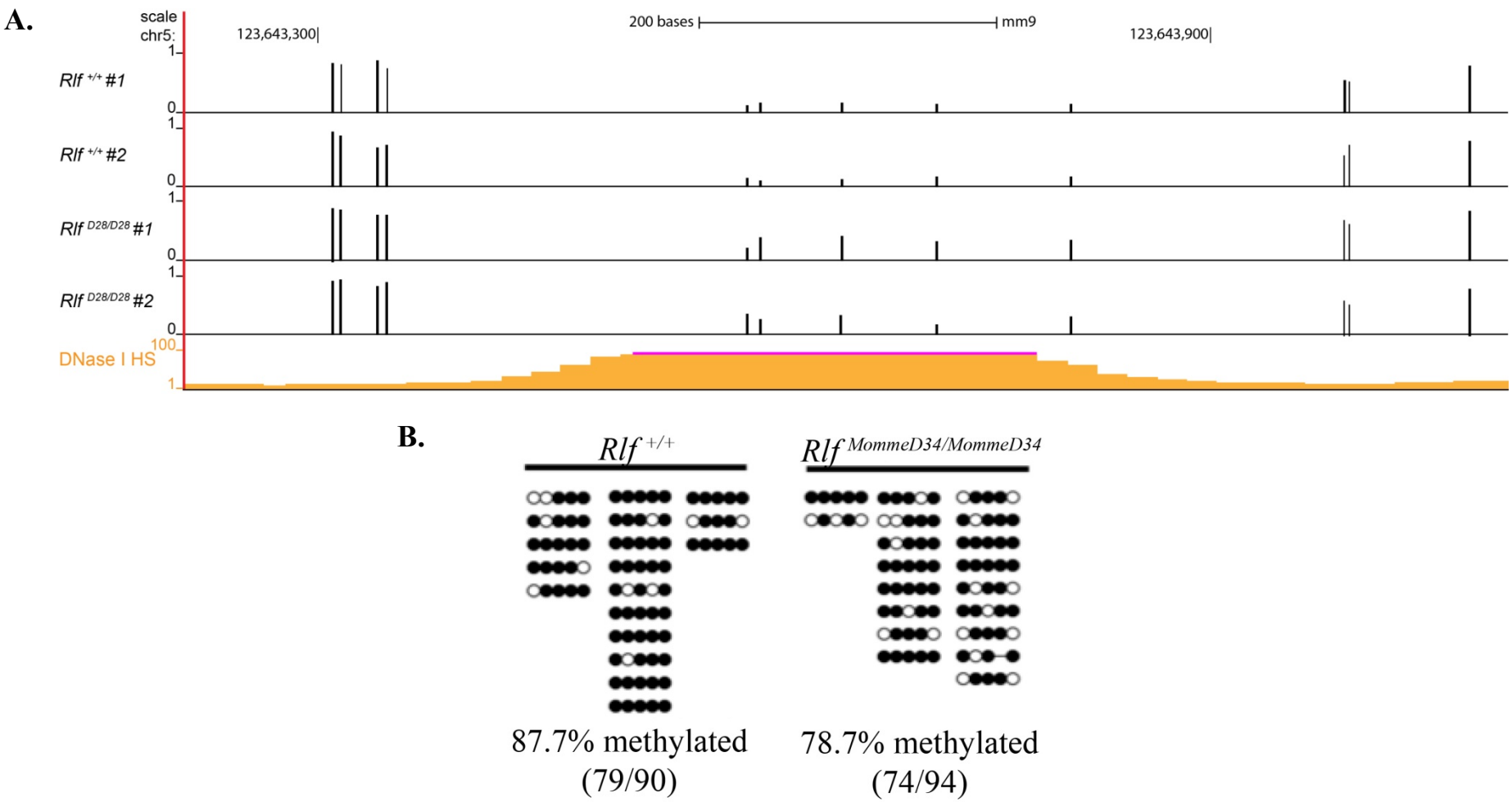
While a ten percent increase in methylation in $Rlf^{MommeD28/MommeD28}$ fetal liver was observed in the GWBS dataset, a decrease was observed following bisulphite sequencing validation in $Rlf^{MommeD34/MommeD34}$ (78%) compared to $Rlf^{+/+}$ (87%) littermates, $p = 0.1489$, **Figure 3.7B**. Possible reasons for these differences in methylation will be discussed in **Section 3.3.** Also of note is the large discrepancy in methylation between wild-types from the *MommeD28* line (~10%) and *MommeD34* line (~87%). One possible explanation for the differences in methylation observed between the two methods may be due to poor read coverage occurring at this region in GWBS data, resulting in methylation at this region being inaccurately defined. Studies investigating the minimum sequencing depth in GWBS have shown short

regions with few CpGs, such as the 5 CpG sites analysed here, are the most susceptible to being missed due to reduced coverage of reads at these regions, resulting in the large different observed between the two methods (Ziller et al., 2015).

Next Page:
Figure 3.7: Methylation analysis of *Hpd* in the fetal liver.

A. UCSC Genome Browser tracks of GWBS results for region chosen for validation upstream of Hpd. The top 4 tracks display samples used in GWBS (n = 2 *Rlf* $^{+/+}$, top, and 2 *Rlf* $^{MommeD28/MommeD28}$, bottom). The DNase I hypersensitivity site (DHS) upstream of Hpd investigated is shown as a yellow track, produced by the University of Washington as part of the publicly available ENCODE mouse project. B. Bisulphite sequencing of this DHS revealed a small decrease in DNA methylation in *Rlf* $^{MommeD34/MommeD34}$ versus *Rlf* $^{+/+}$ fetal liver. Each column represents DNA from a single embryo (n = 3 per genotype), each row is the sequence from a single cell and each circle represents one CpG site. Black circles represent methylated CpGs and white circles represent unmethylated CpGs. (X^2, p = 0.1489).

Figure 3.7: Methylation analysis of *Hpd* in the fetal liver.



A CpG island containing 24 CpGs across the transcription start site of *Prss50* was identified as being differentially methylated in *Rlf* $^{MommeD28/MommeD28}$ (~37%) compared to *Rlf* $^{+/+}$ (~21%) from our GWBS dataset, **Figure 3.8A**. A smaller change in methylation was observed between *Rlf* $^{MommeD34/MommeD34}$ (~25%) and *Rlf* $^{+/+}$ (~20%) fetal liver following bisulphite sequencing, p = 0.4242, **Figure 3.8B**. Comparison of the methylation level in wild-type fetal liver for each of the *Momme* lines also found them to be the same (MommeD34 ~20% and *MommeD28* ~21%). In order to be classified as a Rlf-DMR following GWBS analysis regions were required to have a minimum methylation difference of 15% between the two genotypes (refer next **Section 3.2.5**). Although the *Prss50* region was defined as a Rlf-DMR following GWBS using this cut off (**Appendix IV**), the 15.5% change in methylation between genotypes observed may be attributable to a number of reasons, which will be discussed below.

Failure to validate the *Hpd* and *Prss50* sites may be the result of differences between the *MommeD28* and *MommeD34* lines. Each of these mouse lines was produced from a screen in which a chemical mutagen was used to introduce random point mutations (Blewitt et al., 2005). It has been estimated that the ENU induced mutation rate in mice generated from this screen is 1:1,000,000 bp (Daxinger et al., 2013). Although these mouse lines were backcrossed to unmutagenised FVB/NJ mice, it may be possible that an independent ENU generated mutation is still present in one of the *Momme* lines resulting in differences in methylation between the independent *Momme* lines.

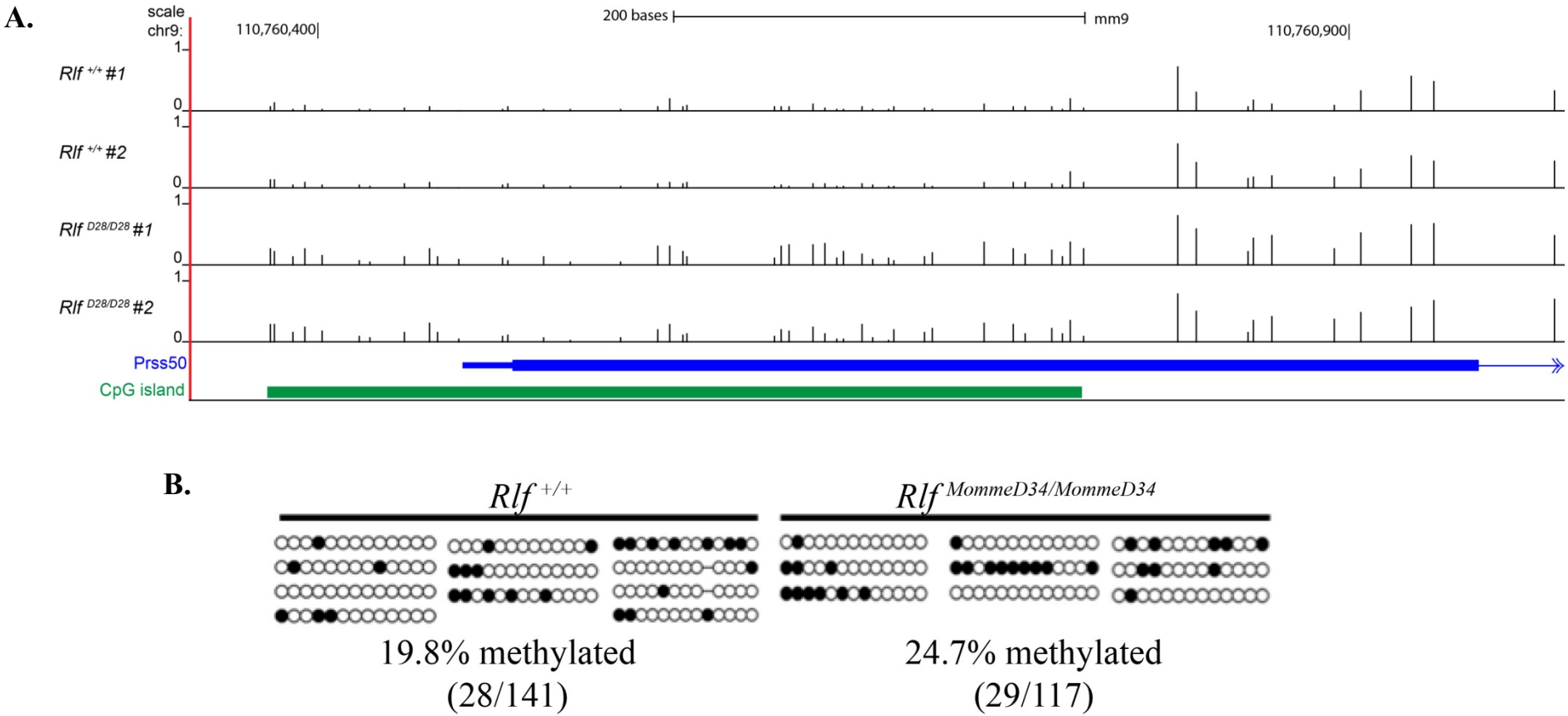
Due to the high costs of performing GWBS on multiple samples, GWBS was only performed in the *MommeD28* mouse line. Undertaking this same analysis in the *MommeD34* mouse line may provide an overlap of differentially methylated regions between the two *Rlf* mutant strains that are less sensitive to heterogeneity between the two lines.

Alternatively, whilst the *Prss50* region was defined as a Rlf-DMR following GWBS analysis, the 15.5% change in methylation was close to the set cut off in methylation change of 15%. The observed 5% increase in methylation in *Rlf* $^{MommeD34/MommeD34}$ mice was in the same direction, but was not as large as the 15% increase observed in *Rlf* $^{MommeD28/MommeD28}$ mice. As such, this change in methylation may be more susceptible to changes in biological variation between replicates.

Next Page:
Figure 3.8: Methylation analysis of *Prss50* in the fetal liver.

A. UCSC Genome Browser tracks of GWBS results for region chosen for validation at Prss50. The top 4 tracks display samples used in GWBS (n = 2 *Rlf* $^{+/+}$, top, and 2 Rlf $^{MommeD28/MommeD28}$, bottom). At the CpG island (green bar) over the first exon of Prss50 (blue bar), an increase in methylation for both mutants is observed in comparison to the wild-type tracks. B. Bisulphite sequencing of the CpG island at the transcription start site of Prss50 displays no difference in methylation. Each column represents DNA from a single embryo (n = 3 for each genotype), each row is the sequence from a single cell and each circle represents one CpG site. Black circles represent methylated CpGs and white circles represent unmethylated CpGs. (X^2, p = 0.4242).

Figure 3.8: Methylation analysis of *Prss50* in the fetal liver.



3.2.5 Identification of Rlf-DMRs in the fetal liver

In parallel with the validation of candidate regions, unbiased identification of Rlf-DMRs using a bioinformatic analysis pipeline was also performed, in collaboration with Dr H. Oey. In these experiments regions were deemed to be significantly differentially methylated between the two genotypes as requiring > 10 CpGs, each CpG must be covered by > 5 reads, and the difference in methylation between the two genotypes be greater than 15% (Harten et al., 2015). These regions were defined as Rlf Differentially Methylated Regions (Rlf-DMRs). 1,329 regions were found to be significantly differentially methylated in *Rlf* $^{MommeD28/MommeD28}$ E14.5 fetal liver compared to *Rlf* $^{+/+}$ following GWBS (**Appendix IV**). **Table 3.7** lists the top 25 Rlf-DMRs and their closest gene/s.

I then graphed the average CpG methylation in *Rlf*$^{MommeD28/MommeD28}$ and *Rlf* $^{+/+}$ littermates and found 94% of Rlf-DMRs were found to be hypermethylated in *Rlf* $^{MommeD28/MommeD28}$, **Figure 3.10**. This result is consistent with previous data showing loss of *Rlf* results in an increase in methylation of the GFP transgene (Daxinger et al., 2013). No difference in the average number of CpGs per DMR was observed when comparing hypermethylated Rlf-DMRs (18.7 CpGs) to hypomethylated Rlf-DMRs (16.7 CpGs, $p = 0.1650$). However, hypermethylated Rlf-DMRs were found to have a significantly greater CpG density (0.055%) than hypomethylated Rlf-DMRs (0.045%, $p = 0.0012$).

Further analysis by Dr H. Oey found approximately half of the 1,329 Rlf-DMRs overlapped with RefSeq transcripts (n=652) and half were intergenic (n=677), **Figure 3.9A**. A small proportion of the Rlf-DMRs, 255, lay within 2.5 kb of a transcriptional start site (TSS) and few, 200, overlapped with the 16,000 CpG islands annotated in the mouse genome. Together these findings suggest that a minority of Rlf-DMRs overlap with promoters (Harten et al., 2015).

Of those Rlf-DMRs that overlapped with RefSeq transcripts, a large proportion, ~50%, were found to be located at exons (including the 3'UTR), **Figure 3.9B**. This is significantly more than would be expected based on the proportion of genic sequence that is exonic (~10% of RefSeq transcripts), **Figure 3.9C** (Harten et al., 2015).

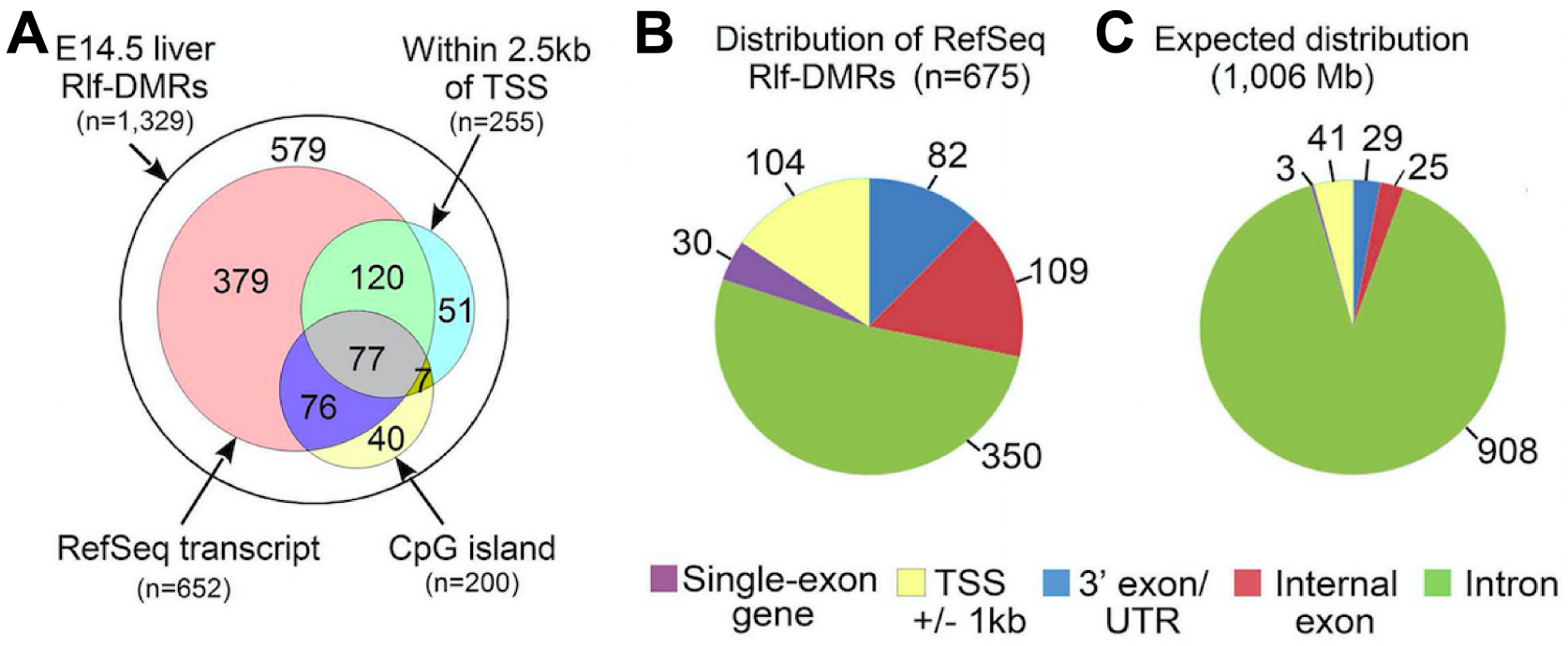


Figure 3.9: Rlf-DMRs overlap with regulatory regions.

A. E14.5 Rlf-DMRs were investigated for overlap with RefSeq genes, proximity to TSS and CpG islands. B. E14.5 liver Rlf-DMRs overlapping RefSeq transcripts, were classified according to overlap with transcript features. The central CpG dinucleotides of each Rlf-DMR was used to define the overlap. Rlf-DMRs that overlapped multiple features were assigned to a single feature according to following ranking: TSS > single exon transcripts > 3' exon/untranslated region (UTR) > internal exon > intron. C. The expected distribution was defined as the union of each feature category genome-wide, subtracting overlapping features of higher rank (1,006 Mb).

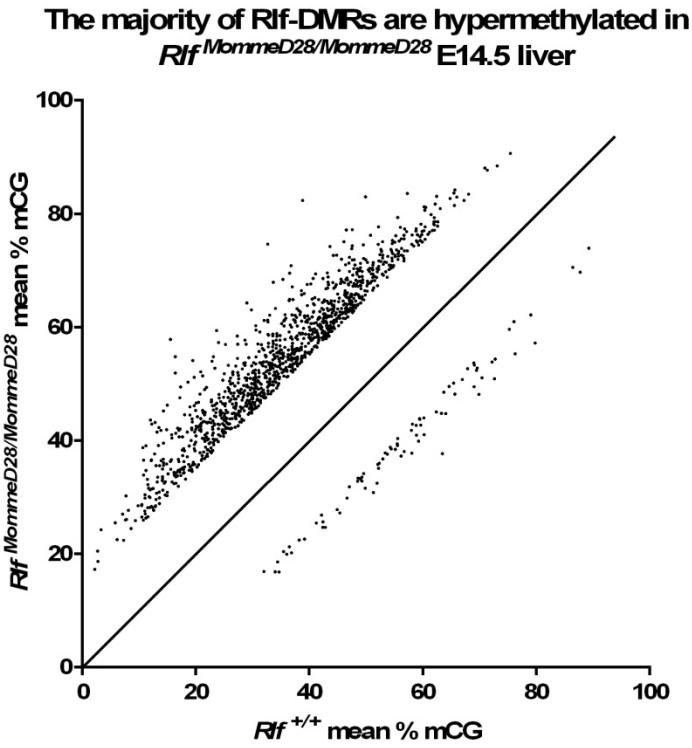


Figure 3.10: The majority of Rlf-DMRs are hypermethylated in *Rlf*$^{MommeD28/MommeD28}$ fetal liver.

Scatterplot of average CpG methylation observed in *Rlf* $^{MommeD28/MommeD28}$ mutants compared to *Rlf* $^{+/+}$ littermates for 1,329 Rlf-DMRs identified in E14.5 liver, shows more Rlf-DMRs are hypermethylated in *Rlf* $^{MommeD28/MommeD28}$ fetal liver than *Rlf* $^{+/+}$ fetal liver.

Next Page:

Table 3.7: List of genes closest to the top 25 Rlf-DMRs in *Rlf* $^{MommeD28/MommeD28}$ fetal liver compared to *Rlf* $^{+/+}$ following GWBS analysis. All of these sites are hypermethylated in *Rlf* $^{MommeD28/MommeD28}$. Fetal liver was collected from two individual E14.5 embryos for each genotype; all embryos were from the same litter. GWBS was performed on each sample individually by CNAG, and bioinformatic analysis was provided by Dr H. Oey. Rlf-DMRs were defined as requiring > 10 CpG sites, each CpG requiring > 5 reads and more than 15% change in methylation between genotypes. Rlf-DMRs are ordered by AreaStat which is equal to the sum of the *t* statistic in each CpG (area of DMR, weighted by the number of CpGs) (Full data Appendix IV).

Table 3.7: Top 25 *Rlf* Differentially Methylated Regions (DMRs) in *Rlf*$^{MommeD28/MommeD28}$ fetal liver

Symbol	Gene Name	CpG sites	Direction	Mean Diff	Area Stat
Rbmxl2	RNA binding motif protein, X-linked-like 2	223	hyper	0.21	818.6
Basp1	Brain abundant, membrane attached signal protein 1	100	hyper	0.38	603.5
Begain	Brain enriched guanylate kinase-associated	120	hyper	0.21	386.7
Hspa1b	Heat shock protein 1B	89	hyper	0.20	361.7
Lrrc4b	Leucine rich repeat containing 4B	98	hyper	0.18	354.0
Irgc1	Immunity related GTPase family, cinema 1	83	hyper	0.24	332.1
Arhgef19	Rho guanine nucleotide exchange factor (gef) 19	56	hyper	0.33	326.3
Jph2	Junctophilin 2	91	hyper	0.19	311.9
Dtnbp1/Mylip	Dystrobrevin binding protein 1/ Myosin regulatory light chain interacting protein	50	hyper	0.35	310.4
Ptger1	Prostaglandin E receptor 4	78	hyper	0.25	305.6
Panx2	Pannexin 2	95	hyper	0.20	304.7
AI854703	Expressed sequence AI854703	68	hyper	0.24	295.5
Nrtn	Neurturin	97	hyper	0.15	293.6
Lingo1	Leucine rich repeat and Ig domain containing 1	95	hyper	0.18	292.6
B3gnt3	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3	102	hyper	0.20	289.7
Zfp575	Zinc finger protein 575	59	hyper	0.29	278.8
Otop3	Otopetrin 3	64	hyper	0.27	266.7
Kcnc1	Potassium voltage gated channel, Shaw-related subfamily, member 1	78	hyper	0.15	262.8
Adra1a	Adrenoreceptor alpha 1A	48	hyper	0.30	255.6
Padi1	Peptidyl arginine deiminase, type 1	39	hyper	0.42	251.2
Smad3/Smad6	SMAD family member 3/ SMAD family member 6	50	hyper	0.26	240.6
Kif21a/Abcd2	Kinesin family member 21A/ ATP-binding cassette, subfamily D (ALD), member 2	38	hyper	0.36	229.6
Mmp9	Matrix metallopeptidase 9	39	hyper	0.31	227.5
Ado/ Zfp365	2-aminoethanethiol (cysteamine) dioxygenase/ Zinc finger protein 365	39	hyper	0.32	221.7
Sod3	Superoxide dismutase 3, extracellular	50	hyper	0.24	221.0

3.2.6 Validation of differentially methylated regions in *MommeD34*

Following the identification of these 1,329 Rlf-DMRs, I asked whether regions identified as significantly differentially methylated in our GWBS dataset could be successfully validated using bisulphite sequencing in an independent mouse line. Two Rlf-DMRs identified following bioinformatic analysis of GWBS data were chosen for bisulphite sequencing validation. The *MommeD34* mutant line was used for validation of *MommeD28* GWBS results, and DNA underwent conversion and sequencing as described previously (refer **Section 3.2.4**). The first region chosen for validation was located between *Smad3* and *Smad6* (Smad family members 3 and 6) on Chromosome 9. This region has been validated as an active enhancer in transgenic mouse assays (VISTA database, www.enhancer.lbl.gov) (Visel et al., 2007). Visel et al., 2007, identified candidate enhancer sequences from evolutionarily conserved sequences or ChIPseq data and cloned these elements into a vector containing a *lacZ* reporter construct (Kothary et al., 1989). The construct was then injected into fertilised oocytes and embryos harvested at E11.5, and stained for *lacZ* reporter activity, and enhancers defined as positive if they showed expression in more than three independent embryos (Pennacchio et al., 2006; Visel et al., 2007). Enhancer hs660, located between *Smad3* and *Smad6* showed expression in the forebrain and neural tube, **Figure 3.11** (Visel et al., 2007).

GWBS analysis of 50 CpG sites in this region found a 26% increase in methylation in $Rlf^{MommeD28/MommeD28}$ fetal livers (53%) compared to wild-type (27%), **Figure 3.11B**. Bisulphite sequencing of 16 CpG sites within this region in the *MommeD34* line revealed a 40%, $p<0.0001$, increase in methylation (87% in

Rlf $^{MommeD34/MommeD34}$ compared to 40% in *Rlf* $^{+/+}$), **Figure 3.11C**. The observed increase in methylation in both homozygous *MommeD28* and *MommeD34* mice, suggests this is truly due to loss of *Rlf* in these mice. The difference in methylation of wild-type controls observed in the two methods and *Rlf* mutant lines may be a result of the region chosen for validation. Primers for bisulphite sequencing validation could only be designed to encompass 16 of the original 50 CpG sites. The high CG content of this region meant specific primers could not be designed to cover the whole region but instead only a small portion, resulting in the 16 CpG sites analysed being skewed to the second half of the differentially methylated region (**Figure 3.11B**). Performing bisulphite sequencing validation of all 50 CpG sites may reveal the level of methylation in wild-type mice from both *Momme* lines to be similar across the whole Rlf-DMR.

Next Page:

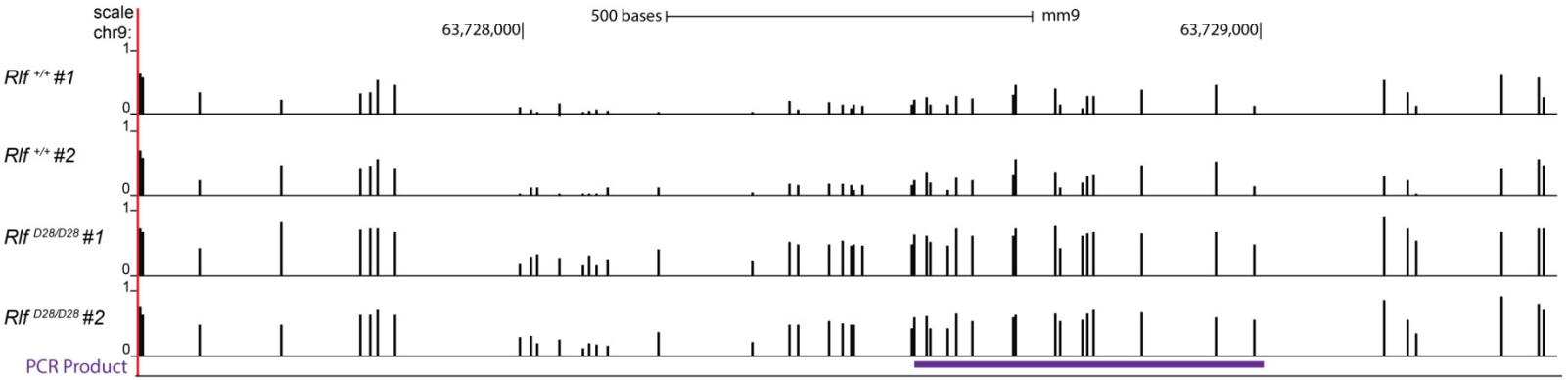
Figure 3.11: Methylation analysis of Smad3/Smad6 region in the fetal liver.
A. *In vivo* reporter assay validating enhancer activity for VISTA enhancer hs660 that overlaps a Rlf-DMR located between *Smad3/Smad6*. Expression is neural tube and forebrain specific, obtained from VISTA enhancer browser (Visel et al., 2007). B. UCSC Genome Browser tracks of GWBS results for regions chosen for validation of *Smad3/Smad6* region. The top 4 tracks display samples used in GWBS (n = 2 *Rlf* $^{+/+}$, top, and 2 *Rlf* $^{MommeD28/MommeD28}$, bottom). An increase in methylation is observed in the mutant versus the wild-type tracks. The PCR region analysed in bisulphite sequencing validation is highlighted below the tracks in purple. C. Bisulphite sequencing of region between *Smad3*/*Smad6* on Chromosome 9 reveals an increase in DNA methylation in *Rlf* $^{MommeD34/MommeD34}$ compared to wild-type fetal liver. Each column represents DNA from a single embryo (n = 3 for each genotype), each row is the sequence from a single cell and each circle represents one CpG site. Black circles represent methylated CpGs, white circles represent unmethylated CpGs, and lines (-) represent an ambiguously sequenced position where a CpG exists in the genomic sequence. (X^2, $p < 0.0001$).

Figure 3.11: Methylation analysis of Smad3/6 region in the fetal liver.

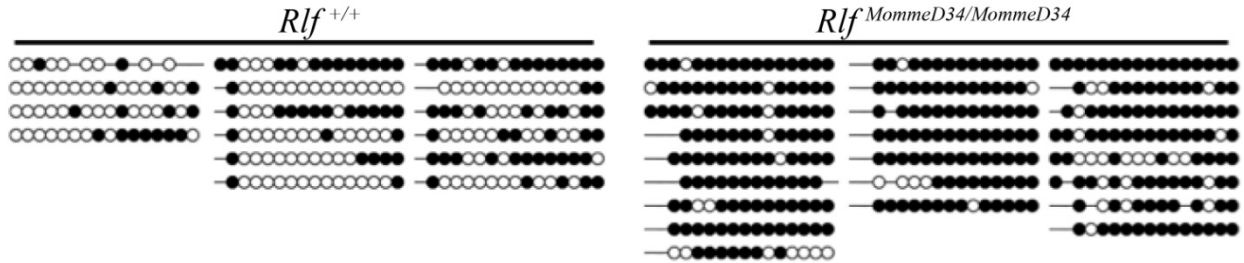
A.

Enhancer Hs660

B.



C.



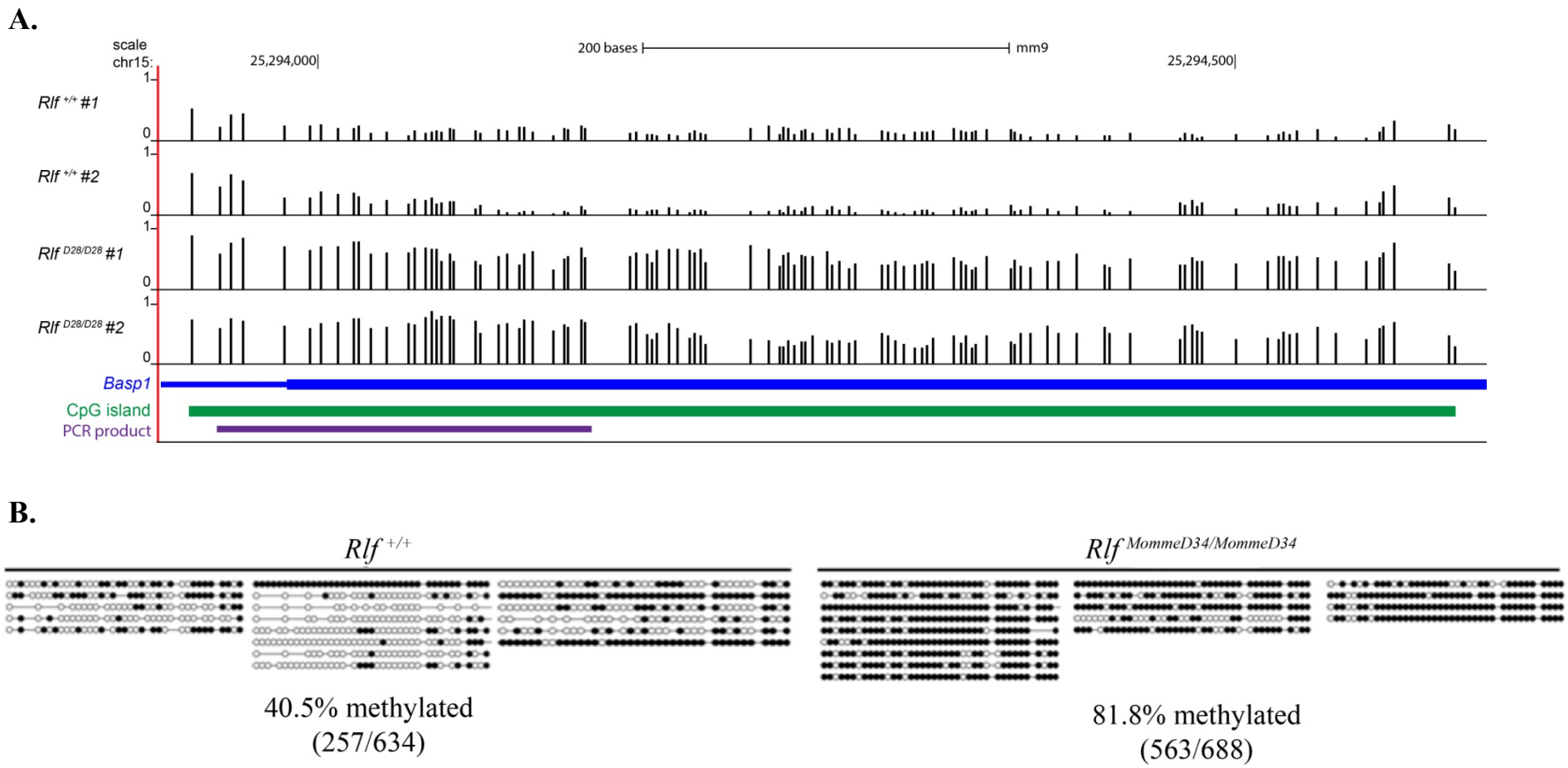
The second region chosen for validation was an intragenic region on Chromosome 15 located over the last exon of *Basp1* (Brain abundant, membrane attached signal protein 1). GWBS analysis of this region found a 38% increase in methylation in *Rlf* $^{MommeD28/MommeD28}$ (55%) compared to wild-type (16%), **Figure 3.12A**. Bisulphite sequencing of 39 of the 100 CpG sites present found a 40% increase, $p<0.0001$, in methylation in *Rlf* $^{MommeD34/MommeD34}$ (82%) fetal liver compared to wild-type (41%), **Figure 3.12B**. This was consistent with Dr Oey's bioinformatics analysis of GWBS data. The validation of GWBS results in an independent *Momme* line confirms that the regions identified are a result of loss of *Rlf*. Furthermore, validation of these two sites gives us confidence that the parameters used for defining Rlf-DMRs are suitable and effective.

Next Page:

Figure 3.12: Methylation analysis of Basp1 region in the fetal liver.

A. UCSC Genome Browser tracks of GWBS results for regions chosen for validation of *Basp1* region. The top 4 tracks display samples used in GWBS (n = 2 *Rlf* $^{+/+}$, top, and 2 *Rlf* $^{MommeD28/MommeD28}$, bottom), an increase in methylation is observed in the mutant versus the wild-type tracks. The CpG island (green bar) over the last exon of Basp1 (blue bar) is shown below GWBS tracks, as well as the PCR region analysed in bisulphite sequencing validation (purple). B. Bisulphite sequencing of *Basp1* region on Chromosome 15 reveals an increase in DNA methylation in *Rlf* $^{MommeD34/MommeD34}$ compared to wild-type fetal liver. Each column represents DNA from a single embryo (n = 3 for each genotype), each row is the sequence from a single cell and each circle represents one CpG site. Black circles represent methylated CpGs, white circles represent unmethylated CpGs, and lines (-) represents an ambiguously sequenced position where a CpG exists in the genomic sequence. (X^2, $p < 0.0001$).

Figure 3.12: Methylation analysis of Basp1 region in the fetal liver.



3.2.7 Rlf-DMRs may reflect active enhancers in early development or different tissues

Following the successful validation of two Rlf-DMRs in an independent mouse lines, and the observation that one of these overlapped a validated enhancer, I wanted to determine whether other Rlf-DMRs also overlapped regulatory regions in the genome. Initial characterisation of the 1,329 Rlf-DMRs by Dr H Oey, found them to be short (~ 1kb) and occurring at regions of the genome that are less methylated than the surrounding DNA (Harten et al., 2015). These features are reminiscent of recently identified tissue specific DMRs (tsDMRs) (Hon et al., 2013a). tsDMRs are defined as regulatory elements thought to be involved in tissue specific differentiation. Some tsDMRs were found to mark enhancers that are dormant in adult tissue but active during embryonic development. Hon et al., 2013, termed these 'vestigial enhancers' and noted they were hypomethylated and lacking active histone marks in adult tissue, but were likely to have been active at earlier stages in development.

Given the similarities between Rlf-DMRs and tsDMRs I asked whether any DMRs were consistent between the two groups (bioinformatics was performed by Dr H. Oey and data interpreted by myself). Fifty-nine percent of Rlf-DMRs were found to overlap with a tsDMR (data not shown). As some tsDMRs overlap with active histone marks present in early development but not in adults, I then wanted to determine whether Rlf-DMRs overlapped with active histone marks in the fetal liver. Our data was compared to publicly available ENCODE E14.5 liver ChIPseq datasets, to determine whether Rlf-DMRs overlapped with histone marks associated with gene activation (The ENCODE Project Consortium, 2012). The histone marks investigated

were H3K4me1 (associated with enhancers), H3K27ac (distinguishes active enhancers from inactive enhancers) and H3K4me3 (associated with active promoters). Comparison of the Rlf-DMRs to ENCODE data by Dr H. Oey found 45 of the 1,329 Rlf-DMRs were enriched for each of the three histone marks analysed, **Figure 3.13A**. Furthermore, 287 were enriched for the H3K4me1 mark alone, 140 for H3K27ac, and 130 for H3K4me3. However, the majority of the fetal liver Rlf-DMRs (973 of the 1,329) did not overlap any of these histone marks, **Figure 3.13A**. One explanation for this could be that Rlf-DMRs may be active enhancers in another tissue or at a different stage of development.

To address this I asked whether Rlf-DMRs overlap with active histone marks in other tissues. Our E14.5 fetal liver data was compared to ENCODE's H3K4me1 ChIPseq data in E14.5 brain and heart. An additional 478 Rlf-DMRs were enriched for the enhancer mark H3K4me1 in one or both of these tissues, **Figure 3.13B.** These results suggest that many, if not all, of the Rlf-DMRs without active histone marks in E14.5 liver may display enhancer marks in other tissues, one of which will be analysed in **Section 4.2.16.**

Gene ontology analysis of the genes closest to the 1,329 Rlf-DMRs did not identify a specific class of gene except for an enrichment of genes encoding for transcription factors (Dr. H Oey, *personal communication*) (Harten et al., 2015). Consensus DNA binding motifs at Rlf-DMRs were also investigated with no motifs found to be enriched significantly over the background (Dr. H Oey, *personal communication*) (Harten et al., 2015).

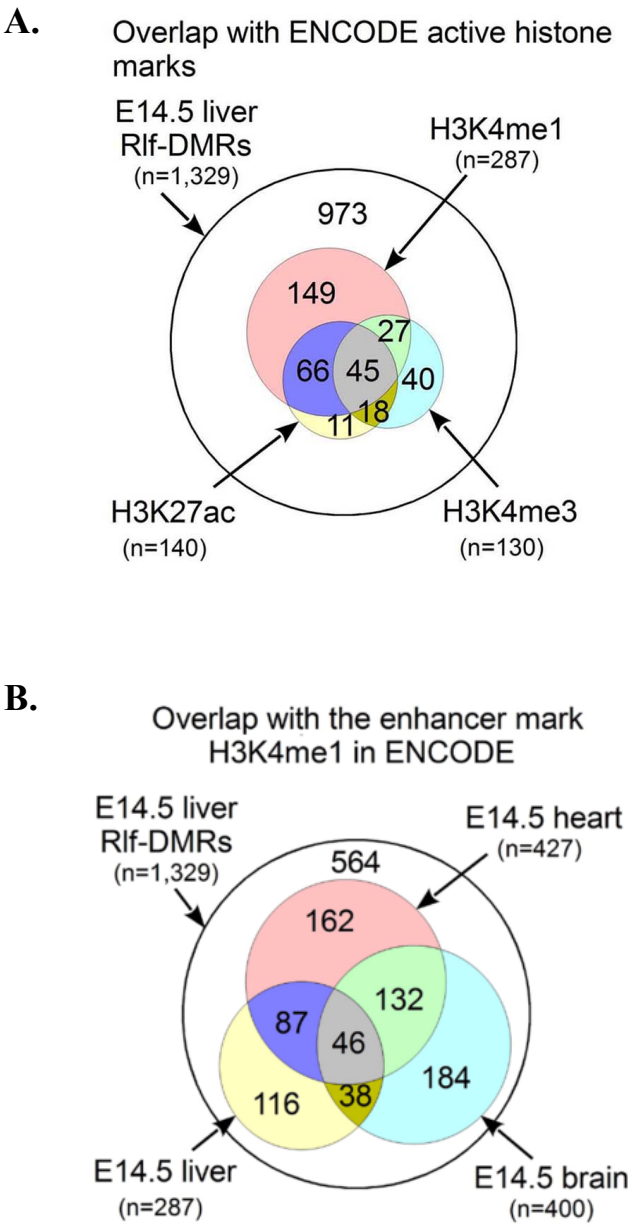


Figure 3.13: Rlf-DMRs overlap with active histone marks identified in ENCODE E14.5 liver, heart and brain tissue.

A. Overlap of E14.5 fetal liver Rlf-DMRs with the histone marks H3K4me1, H3K4me3 and H3K27ac in E14.5 liver (obtained from ENCODE). B. Overlap of E14.5 fetal liver Rlf-DMRs with the enhancer mark H3K4me1 in ENCODE E14.5 liver, heart and brain tissues. ChIPseq enrichment was defined as elements with ChIPseq Reads Per Kilobase per Million mapped (RPKM) greater than two-fold enriched over input RPKM for both replicates. Figure produced by H. Oey and published (Harten et al., 2015).

3.2.8 Chromatin immunoprecipitation sequencing of Rlf fetal liver

The RNA-seq and GWBS experiments undertaken in this chapter do not indicate whether changes observed in the methylome or transcriptome are a direct result of the presence of Rlf binding at these sites. Previous work by our laboratory has found that some, but not all of the zinc fingers present in Rlf are capable of directly binding DNA (Harten et al., 2015). As a result of this I wanted to determine where in the genome Rlf may bind using ChIPseq, which allows for genome-wide identifications of putative chromatin occupancy sites (Zhang et al., 2012b).

Two independent Rlf antibodies designed to different termini of Rlf were used for ChIPseq analysis. The use of antibodies that that recognise different regions of Rlf is recommended by the ENCODE consortia for characterising antibody binding and is expected to be beneficial in defining the specificity of the Rlf antibodies (Landt et al., 2012). The commercial Abcam Rlf antibody (AB115011) is designed against the C-terminal end of human RLF (mouse amino acid 1845 – 1918), with 93% homology to mouse Rlf, whilst the custom designed Rlf antibody, Ab1 (please refer **Section 4.2.1**), is designed against the N-terminal end of mouse Rlf (amino acid 265-279) (refer **Figure 4.1**). Making these antibodies suitable, according to ENCODE recommendations, for ChIPseq identification of putative Rlf binding sites.

Fetal liver was collected from eight *Rlf* $^{+/+}$ E14.5 embryos that were then divided into four replicates, consisting of two livers in each replicate. Fetal liver was

then sent to ActiveMotif for chromatin extraction and immunoprecipitation using either the commercial Abcam Rlf antibody custom Rlf antibody Ab1.

Sequencing of Abcam samples and their input control was undertaken on an Illumina HiSeq 2000 with 50 nt reads, whilst the Ab1 samples and their input control were sequenced on an Illumina NextSeq 500 with 75 nt reads. Sequencing was performed by ActiveMotif and all of the following bioinformatics analysis was undertaken by myself. All reads were mapped to the mouse genome (NCBI37/mm9) using the Burrows-Wheeler Aligner (BWA) algorithm with default settings (Li & Durbin, 2009). Reads passing Illuminas purity filter, aligned with no more than two mismatches and map uniquely to the genome were used in subsequent analyses. Samples from each experiment were normalised alongside of their respective input controls to the same number of unique alignments following the removal of duplicate reads (**Table 3.8**).

Peaks were identified using model based analysis for ChIPseq (MACS v1.4.2) (Zhang et al., 2008) peak finding algorithm with a cut off *p* value of $p = 1e\text{-}7$ (Ab1) and $p = 1e\text{-}5$ (Abcam). Following ChIPseq with the Abcam antibody a different *p* value cut off was used due to the low number of unique alignments after the removal of duplicate reads (6.5 million, **Table 3.8**). This low number of unique alignments indicates there was low coverage in these samples which may have been the result of inadequate starting material or excessive amplification in the PCR step (Bailey et al., 2013). In comparison the number of unique reads following the removal of duplicates for the Ab1 analysis was approximately three times greater (18.3 million,

Table 3.8), indicating a much higher coverage in these samples. Additionally, experiments analysing ChIPseq data of histone modifiers have identified broad peaks, spanning larger regions, that are less enriched than control samples, making less stringent *p* value cut offs more useful in initial analysis (Feng et al., 2012). As it is unknown where in the genome Rlf may bind or whether it is enriched across large or short regions, the less stringent cut off for this experiment was chosen.

Each of the replicates that underwent ChIPseq with the Abcam antibody were found to have over 3,800 peaks, of which a small number were found to be on the ENCODE Blacklist for false-positive ChIPseq peaks (~100 in each sample **Table 3.9**) and were excluded from further analysis. Following the removal of these false positive peaks over 3,700 peaks remained in each sample, 1,681 of which were found in both replicates, **Table 3.9** (full data **Appendix V**).

From the replicates that underwent ChIPseq with the custom designed Rlf antibody, one sample (Wild 5) was found to have > 9,000 peaks, whilst only 409 peaks were identified in the other sample (Wild 6), **Table 3.9** (Full data **Appendix VI**). No peaks were found to be on the ENCODE Blacklist for false-positive peaks in either sample and the false discovery rate was zero. Ninety-five percent of the peaks identified in Wild 6 overlapped with those identified in Wild 5, resulting in 381 peaks called in both replicates.

Comparison of results from the two independent ChIPseq experiments identified 202 putative Rlf binding sites that were in each of the biological replicates.

Enrichment at these sites was observed to be higher in the Ab1 Wild 5 replicate than all other replicates, from both ChIPseq experiments. This may be indicative of reduced DNA pulled down during immunoprecipitation for the other replicates. A representative Rlf ChIP peak showing differences in enrichment between the replicates and ChIPseq experiments is shown in **Figure 3.14**.

Table 3.8: Sequencing Statistics following ChIPseq with Abcam and Ab1 Rlf antibodies

	Abcam Rlf Antibody			Custom Rlf Antibody		
	Wild 1	Wild 2	Input	Wild 5	Wild 6	Input
Total number of reads	27,771,300	21,328,600	35,982,257	29,966,334	29,782,883	39,935,321
Total number of alignments (mm9)	24,520,594	19,867,160	34,726,380	24,244,952	24,068,296	32,405,757
Unique alignments (-q 25)	19,313,752	15,800,299	27,016,755	20,629,207	20,393,786	26,981,092
Unique alignments (without duplicate reads)	6,523,873	8,054,275	19,495,616	18,365,494	19,080,384	24,973,129
Normalized tags	6,523,873	6,523,873	6,523,873	18,358,881	18,358,881	18,358,881

Table 3.8: Sequencing statistics showing total number of reads for each sample, number of alignments for each sample, unique alignments, and number of alignments following removal of duplicate reads.

Table 3.9: Number of Peaks identified following MACS analysis of Abcam and Ab1 Rlf ChIPseq

	Abcam Rlf Antibody		Custom Rlf Antibody	
	Wild 1	**Wild 2**	**Wild 5**	**Wild 6**
Paired Peaks	1,195	1,219	1,859	308
Final MACs peaks	3,984	3,835	9,005	409
Predicted Fragment length	346	332	270	85
Negative peaks	4	4	0	0
	0.10%	0.10%	0.00%	0.00%
mm9 blacklisted	108	116	0	0
filtered	3,876	3,719	0	0

Table 3.9: Peaks identified following MACs analysis of Abcam and Ab1 Rlf ChIPseq data, showing the number of peaks in each sample and the final number of peaks following the removal of false-positives. (Full list of peaks and genome co-ordinates Appendix V (Abcam) and VI (Ab1)).

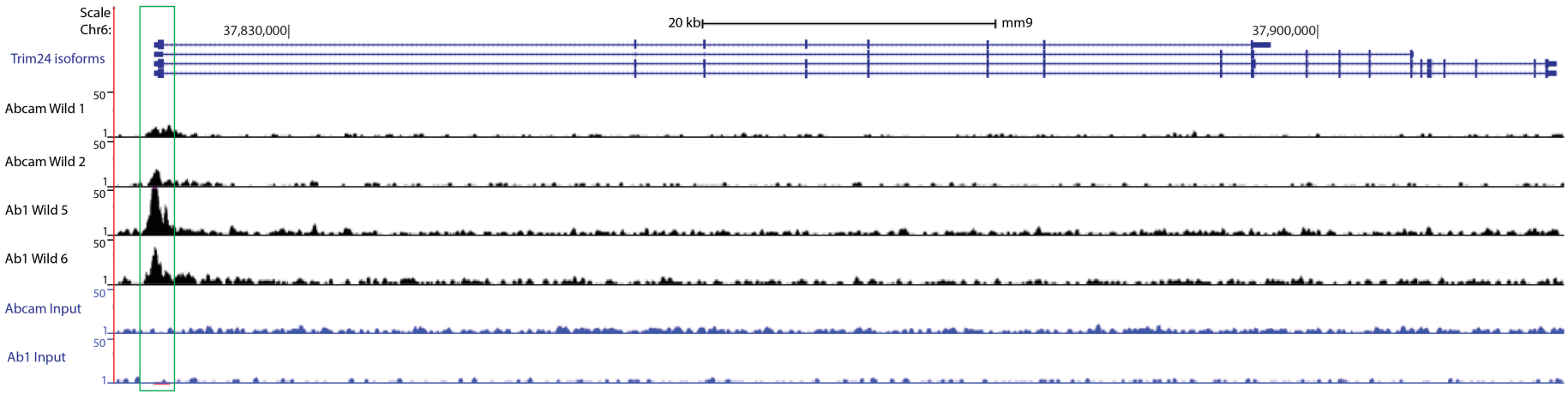


Figure 3.14: Representative ChIP peak identified in all four replicates from ChIPseq experiments with two independent Rlf antibodies.

UCSC Genome Browser tracks of putative Rlf binding site at the transcription start site of *Trim24*, Tripartite motif-containing 24, (blue bars). The top four tracks display enrichment in each of the wild-type replicates used for ChIPseq (Abcam Wild 1 and 2, and Ab1 Wild 5 and 6), and the ChIP peak is highlighted (green box). The respective input controls tracks for each experiment are displayed in the bottom two tracks (blue). Greater enrichment at the putative binding site was observed in the Ab1 Wild5 replicate compared to the other replicates.

Following the identification of these 202 putative Rlf binding sites, I wanted to annotate where in the genome these sites may occur. To do this I used the web-based tool Peak Annotation and Visualisation (PAVIS) to annotate the Rlf ChIP peaks (Huang et al., 2013). Sixty-nine Rlf ChIP peaks were found to be intronic, whilst 45 peaks were located in a 5' untranslated region and 40 peaks within 2.5 kb of a TSS, **Figure 3.15**. The observation of these putative Rlf binding sites close to TSS correlates with Rlf playing a role in transcriptional activation.

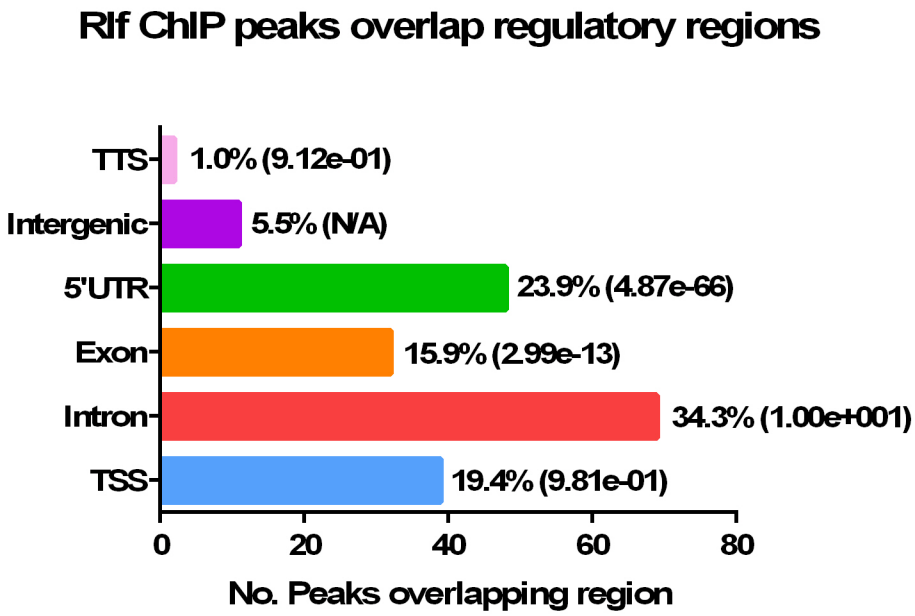


Figure 3.15: Rlf ChIP peaks overlap regulatory regions

ChIP peaks identified using two different Rlf antibodies are classified according to their overlap with transcript features, the percentage and enrichment *p* value as calculated by PAVIS are provided.

I then wanted to determine whether any of these peaks overlapped with genes previously identified as being differentially expressed within the fetal liver. One caveat of using these 202 peaks is that some true Rlf binding sites may have been removed due to low enrichment resulting, in them not being identified in one or more replicates as a peak by MACS peak calling software (Yang et al., 2014). This is important in the case of the Wild 6 biological replicate in which a lower number of peaks were identified compared to all other biological replicates. This may have resulted in a number of 'true/good' peaks being discarded prior to further analysis (Landt et al., 2012). By choosing to analyse these 202 peaks present in all biological replicates across the two ChIPseq experiments, the likelihood of including peaks that are irreproducible, false or representative of noise has been reduced.

Twenty-seven genes that were identified as differentially expressed ($p < 0.05$, FC > 1.2) in *Rlf* $^{MommeD28/MommeD28}$ E14.5 liver were found to have a Rlf ChIP peak within 50 kb of their TSS (**Figure 3.16**). The majority of these ChIP peaks were proximal to genes that were down-regulated ($n = 17$) rather than up-regulated ($n = 10$), consistent with Rlf being required for transcriptional activation. Seven of these genes also had a Rlf-DMR within 50 kb, whilst an additional 50 genes had a Rlf-DMR within 50 kb but no Rlf ChIP peak (**Table 3.10**, **Figure 3.16**). Repeating this analysis to find Rlf ChIP peaks and DMRs within 1 kb of the TSS of differentially expressed genes identified eleven genes with a ChIP peak and an additional two genes with both a ChIP peak and Rlf-DMR (*Dnajb3*, *Mesdc*, data not shown). Investigation of a random gene list of the same size (881 genes total) identified one gene that had both a ChIP peak and Rlf-DMR within 50 kb of its TSS (data not shown). None of the random genes had a Rlf ChIP peak and Rlf-DMR within 1 kb of

its TSS (data not shown). This suggests significantly more differentially expressed genes were found to have Rlf ChIP peaks than random (Fishers exact test, p <0.0001), and the peaks identified are true Rlf binding sites and are not false positives.

I then wanted to determine whether putative Rlf binding sites overlapped with Rlf-DMRs identified in the fetal liver. Of the 202 peaks identified from both ChIPseq experiments, four Rlf-DMRs were found to be within 1 kb of a putative Rlf binding site; whilst forty Rlf-DMRs were found to be within 50 kb of a putative Rlf binding site (data not shown). Repeating this analysis using the two ChIPseq experiments as independent datasets identified 32 Abcam ChIP peaks (out of total 1,681), and eight Rlf Ab1 ChIP peaks (out of total 381) within 1 kb of a Rlf-DMR (data not shown). As few Rlf-DMRs and putative Rlf binding sites were found to overlap this suggests Rlf binding is not the cause of the change in methylation observed. Instead the presence of Rlf binding sites proximal to TSS may suggest that Rlf is stabilising or facilitating an interaction with a complex, through chromatin looping, that is bound to the Rlf-DMR and altering methylation. This will be discussed further in **Section 3.3.**

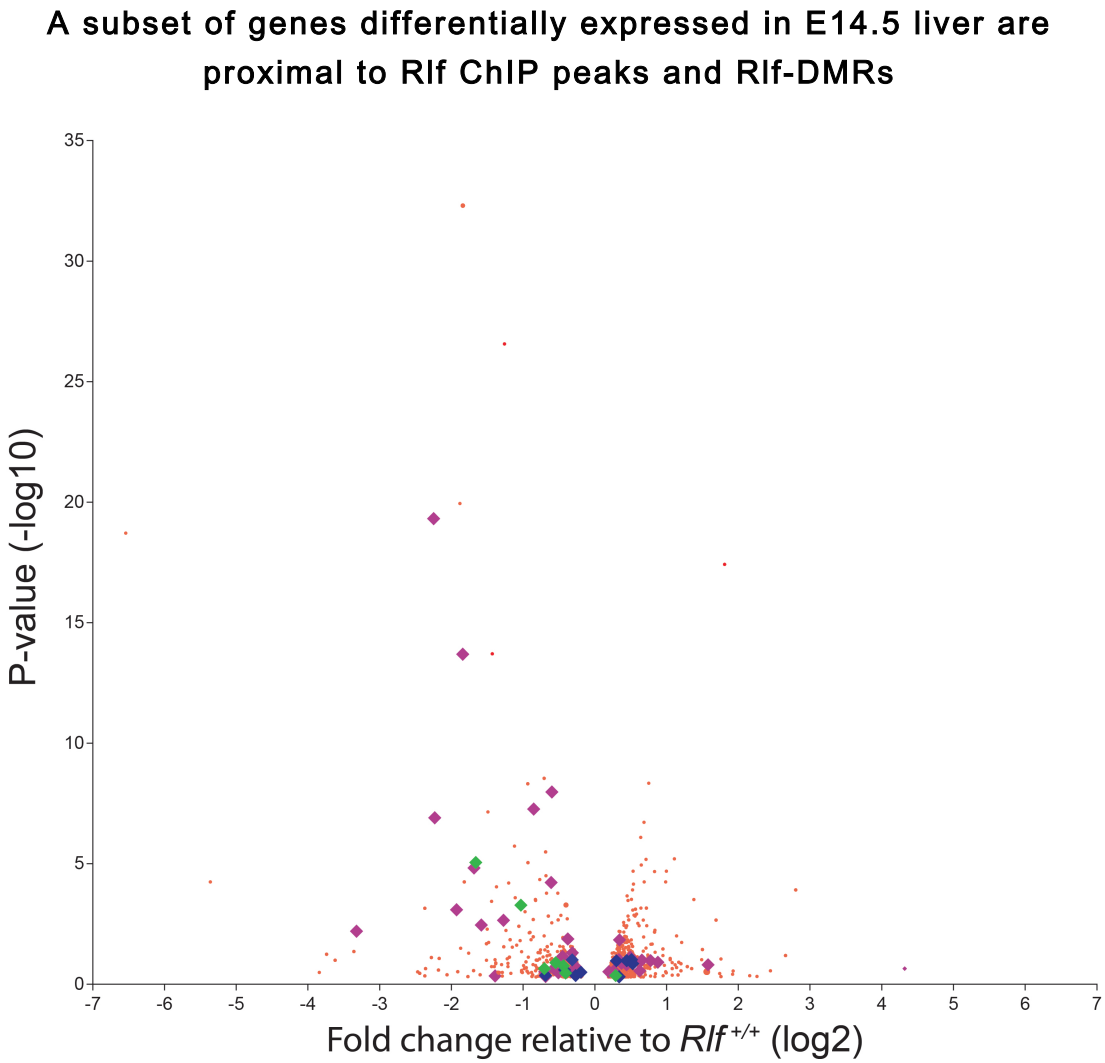


Figure 3.16: A subset of genes differentially expressed in E14.5 liver are proximal to Rlf ChIP peaks and Rlf-DMRs

Volcano plot depicting differentially expressed genes comparing RNA-seq analysis of E14.5 livers from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ mice (n = 3 per genotype). Significantly differentially expressed genes, with an adjusted p value < 0.05, and fold change > 1.2, are presented as red, blue, green or purple data points. Genes that have either an Rlf-DMR or an Rlf-dependent ChIP peak within 50 kb of their TSS are highlighted by purple and blue diamonds, respectively. Green diamonds highlight genes with a change in both DNA methylation and an Rlf-dependent ChIP peak.

Table 3.10: Differentially expressed genes that have both a putative Rlf binding site and an Rlf DMR within 50 kb of their TSS.

Gene Symbol		Rlf ChIP peak co-ordinates		Rlf-DMR Co-ordinates		RNA seq statistics	
	Chr	Start	End	Start	End	Log 2 Fold Change	Adjusted P value
Dnajb3	1	90,051,665	90,115,572	90,101,665	90,102,005	3.162	8.76E-06
Ppapcd3	2	31,951,048	31,966,340	31,951,068	31,951,529	1.63	2.24E-07
0610031J06Rik	3	88,128,945	88,135,235	88,169,691	88,170,207	0.80	4.77E-01
Txn1	4	57,956,245	57,969,283	58,047,491	58,047,890	1.21	3.33E-01
Atf7ip	6	136,467,372	136,518,358	136,503,640	136,504,622	1.20	1.97E-01
Slc27a5	7	13,573,695	13,583,541	13,534,577	13,535,644	2.04	5.19E-04
Mesdc	7	91,029,005	91,032,851	91,030,912	91,031,416	1.22	1.64E-01

Table 3.10: Seven genes identified as being differentially expressed in E14.5 liver following RNA-seq analysis were found to have an Rlf ChIP peak and Rlf-DMR within 50 kb. The chromosome co-ordinates from both the ChIP peak and Rlf-DMR analysis are listed alongside of the genes fold change in expression (in *Rlf* $^{MommeD28/MommeD28}$ mice) and adjusted *p* value determined by RNA-seq analysis.

While a number of putative Rlf binding sites have been identified here, it is important to note the technical limitations that can impact upon ChIPseq results. Antibody quality has been noted as an important factor in the literature. Studies have shown that whilst the majority of antibodies are suitable for ChIPseq, approximately 25% have problems with their specificity (Elgin et al., 2011). Western blotting for Rlf using each of the independent Rlf antibodies investigated here, has found only one band to be present at the expected size for Rlf in the fetal liver (**Figure 3.1** and Harten et al., 2015). However, multiple bands in other tissue types have been observed with these antibodies, and this in turn may impact upon the sensitivity and specificity of these Rlf antibodies in ChIPseq experiments.

Another limitation is the use of suitable controls in ChIPseq analysis. The analysis undertaken here utilised DNA input controls in the absence of *Rlf* null fetal liver DNA. The use of total sheared DNA as an input control is common in ChIPseq experiments and is recommended as by the ENCODE consortia (Landt et al., 2012). Repeating ChIPseq analysis or undertaking validation using ChIP-qPCR with *Rlf* null fetal liver controls would provide confirmation that the putative sites identified here are not due to non-specific antibody binding and are true Rlf binding sites.

3.3 DISCUSSION

In this chapter I have shown for the first time that loss of Rlf alters the transcriptome and methylome in the mouse fetal liver. Previous studies have suggested that Rlf is an enhancer of variegation and the data presented here correlates with these previous findings (Daxinger et al., 2013). In this study I undertook RNA-seq in the *MommeD28* mouse line and validated results using the independent mouse line, *MommeD34*. More genes were found to be down-regulated (n = 51) than up-regulated (n = 31) (**Figure 3.A**) and also the majority of DMRs were hypermethylated (**Table 3.7**) in $Rlf^{MommeD28MommeD28}$ mice, consistent with Rlf being required for transcriptional activation.

One of the genes identified as being markedly reduced in *Rlf* mutant livers was *Hpd*, which plays a key role in the tyrosine catabolism pathway (**Figure 3.4B**) (Tanaka et al., 2006). Published studies show mutant mice null for *Hpd* resulting in no protein being produced have elevated tyrosine levels in the blood and no liver damage (Endo et al., 1991). Histological analysis of *Rlf* mutant compared to wild-type mice also showed no liver damage was present (**Figure 3.1A**). However, we did not observe any increases in blood tyrosine levels in adult *Rlf* mutant mice (**Figure 3.5A**). Interestingly the blood tyrosine levels observed in our adult compound heterozygous mice (63 - 120 μmol/L) were similar to that observed in control mice (50 – 150 μmol/L) in Endo et al., 1991, whilst $Hpd^{-/-}$ mice had approximately 10 fold increase in blood tyrosine (717 – 1,052 μmol/L). As $Rlf^{MommeD8/MommeD34}$ mice still express some Rlf protein, unlike $Rlf^{MommeD34/MommeD34}$ and $Rlf^{MommeD28/MommeD28}$ which produce no Rlf protein (Daxinger et al., 2013), it is

likely that the amount of Hpd produced in our compound heterozygous adult mice (*Rlf* $^{MommeD8/MommeD34}$) is sufficient for catalysing 4-hydroxyphenylpyruvate to homogentisate.

The impact of loss of Rlf on tyrosine in the embryo was also investigated. During human fetal development the accumulation of available tyrosine decreases, late gestation babies were found to have decreased tyrosine accumulation compared with early gestational babies (Polin et al., 2004). This decrease in tyrosine accumulation correlates with independent studies showing the activity of Hpd increases towards late stage gestation (Fellman et al., 1972). Here I have analysed tyrosine serum levels using embryos from the *MommeD34* mouse line. Due to a marked decrease in *Hpd* mRNA expression (**Figure 3.**) an increase in tyrosine was anticipated. A confounding factor in this analysis is *Rlf* $^{MommeD34/MommeD34}$ pups are smaller in size (Daxinger et al., 2013), as potential delays in development may also result in higher tyrosine concentration in mutant mice. However, a decrease in tyrosine was observed in homozygous offspring (**Figure 3.6B**). A possible explanation for our results could be that fetal uptake of tyrosine from the maternal supply is low. Whilst the mothers primary tyrosine source is dietary based, the embryo may instead rely upon endogenous production of tyrosine and the small size of our mutant embryos results in less tyrosine production (Chien, 1991). Additionally, an alternate pathway resulting in the degradation of tyrosine may also be responsible. Further studies would be required to reveal the underlying mechanism in *Rlf* mutant mice.

Previously, the Whitelaw laboratory has shown an increase in methylation of the GFP transgene in mice homozygous for each of the *Rlf* mutant alleles *MommeD8, MommeD28* and *MommeD34*, consistent with a decrease in the number of GFP expressing cells in mutant mice. As changes in transcription can also correlate with changes in methylation we decided to analyse methylation at two genes, *Hpd* and *Prss50*, which were found to be differentially expressed in *MommeD28* mutant mice**.** A DHS site upstream of *Hpd* (**Figure 3.7A**) was analysed as these sites commonly mark enhancers, which can regulate genes that are both proximal and distal to the DHS (Heintzman et al., 2009; Li et al., 2012; Zhang et al., 2012a). Once cut offs were assigned the change in methylation at the *Hpd* region chosen was not considered to be substantial enough to be classified as an Rlf-DMR. As Rlf-DMRs were found to overlap with enhancer marks, which can act over long ranges, a region distal to *Hpd* may be controlling the mRNA expression of this gene. Making the investigation of this DHS site unsuitable for determining whether there is a correlation between methylation and gene expression of *Hpd*.

Unlike the DHS site upstream of *Hpd*, which was not classified as an Rlf-DMR, Rlf-DMRs were identified at CpG islands at the promoters of *Prss50* and *Basp1*. The majority of promoters (72%) have been found to have a high CpG content, while 28% have CpG content that more closely resembles the overall genome CpG content. Genes that have CpG islands across their promoters have also been shown to be expressed in a number of different tissues (Saxonov et al., 2006). Interestingly, these promoters are mostly free of methylation, this correlates with our observation of low methylation of the CpG islands at the promoters of *Prss50* and *Basp1* (**Figure 3.8** and **Figure 3.12**). GWBS revealed an increase in methylation

across both the *Prss50* and *Basp1* promoters, 15% and 40% respectively, and both were classified as a Rlf-DMR following GWBS analysis in the *MommeD28* line. However, only the *Basp1* region was successfully validated using bisulphite sequencing in the *MommeD34* line. To be identified as a differentially methylated region Rlf-DMRs were required to have a greater than 15% change in methylation between the two genotypes. The *Prss50* Rlf-DMR identified in GWBS was only just above this cut off with a 15.5% difference in methylation identified. As our GWBS was undertaken with two replicates per genotype, changes in methylation that are close to the fixed cut off may be difficult to validate successfully. For example, differences in biological variation between two replicates can have a greater influence on the efficiency of statistical testing, than if additional replicates are analysed. Making large changes such as those observed at *Basp1* easier to validate. Alternatively, undertaking validation of the *Prss50* site in the *MommeD28* line, the same mouse line in which GWBS was completed, may uncover this change in methylation as being specific to the *MommeD28* mutation and not the *MommeD34* mutation.

Recent studies investigating differential methylation in different tissues in both mice and humans have found that tsDMRs overlap transcription factor binding sites and histone marks associated with active enhancers (H3K4me1 and H3K27ac) in at least one of the tissues or time-points studied. Hon et al., 2013, found that in late gestation and adult tissue some of these regions lacked active histone marks that were present early in gestation but retained a hypomethylated state. This suggests these hypomethylated regions represent regions that were active earlier in development. The Rlf-DMR we identified and validated between *Smad3* and *Smad6*

in the fetal liver is an example of an Rlf-DMR without active histone marks (**Figure 3.11B**). Interestingly, this region showed reproducible expression in the neural tube and forebrain (**Figure 3.11A**), and was defined as a positive enhancer at E11.5 (Visel et al., 2007). This could suggest that while some of the Rlf-DMRs were not enriched for active histone marks in the fetal liver, they may retain epigenetic memory from a different developmental time-point or tissue.

No previous work has investigated where in the genome Rlf may bind and whether there is a correlation between Rlf binding, transcription and methylation. Here I have used two independent Rlf antibodies, designed to different termini of Rlf, to determine where in the fetal liver genome Rlf may bind. 202 putative Rlf binding sites were identified and the majority of these were found to be enriched proximal to gene transcription start sites (**Figure 3.15**). A small subset of these Rlf binding sites were also found to be proximal to differentially expressed genes and Rlf-DMRs in the fetal liver. Recent studies have shown DNA binding factors, such as CTCF and REST, can influence DNA methylation; for example, the binding of REST to methylated CpG poor regions leads to demethylation (Feldmann et al., 2013; Stadler et al., 2011). Whilst we do not observe Rlf binding at the majority of Rlf-DMRS the binding of Rlf to the TSS of a gene may be facilitating an interaction with a complex (i.e. factor X) that binds to the Rlf-DMR/distal enhancer leading to demethylation. This may occur through chromatin looping bringing the distal enhancer into contact with the TSS allowing for Rlf to interact with factor X and stabilise its binding, **Figure 3.16.** By stabilising this binding the presence of Rlf may be influencing the loss of methylation at the distal site. In cells in which Rlf is

absent, this interaction and stabilisation cannot occur resulting in the distal region remaining methylated.

Other possibilities as to why we did not observe Rlf binding at the majority of DMRs may include that these DMRs have been falsely identified as Rlf-DMRs, such as the case with the Rlf-DMR identified at *Prss50*, Figure 3.8, which was found to not be differentially methylated following clonal bisulphite sequencing. Some of these differentially methylated regions may also be downstream effects occurring due to loss of Rlf, or that they are false positive DMRs. For example, loss of UHRF1 has been found to lead to decreases in DNA methylation as it is no longer able to recruit DNMT1 (Bostick et al., 2007). Loss of Rlf may lead to alterations in DNA methylation in a similar manner; however the mechanism through which this may occur is unclear.

Recent work by our group has also identified fifty percent of Rlf-DMRs overlap with exons, significantly more than would be expected, ten percent, **Figure 3.9** (Harten et al., 2015). This is consistent with studies revealing many exons act as enhancers, having the ability to influence transcription either at the gene in which they reside or a neighbouring gene (Birnbaum et al., 2012). Fifteen percent of our putative Rlf ChIP peaks were also found to overlap with exons, **Figure 3.15**. Whilst these ChIP peaks did not overlap with DMRs the high stringency cut-offs in defining a ChIP peak from the two independent experiments used here may have removed ChIP peaks that did. Further investigations of putative Rlf binding sites throughout

the genome with an antibody of appropriate quality may allow for the mechanism through which Rlf is acting to be elucidated.

These results support previous findings by or laboratory that Rlf is required for transcriptional activation, however there may be other factors in addition to Rlf that are required for transcription to be affected. As Rlf-DMRs are thought to represent active enhancers in other tissues or from earlier stages in development, future experiments investigating the impact of Rlf on chromosome conformation in different tissues and/or developmental time-points may reveal how Rlf acts as an epigenetic modifier.

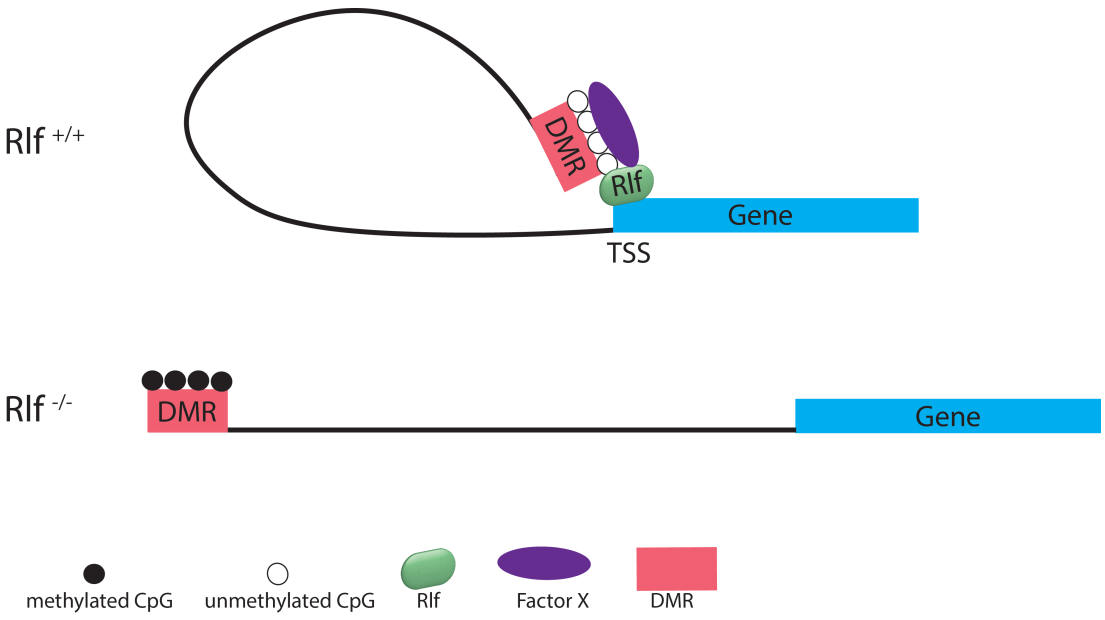


Figure 3.17: Proposed mechanism of Rlf

Proposed mechanism by which Rlf alters transcription. Rlf binding of TSS stabilises an interaction with a complex (factor X) that binds to distal enhancer and leads to demethylation of the distal enhancers.

3.4 FUTURE DIRECTIONS

As a large number of epigenetic modifiers act within complexes, and Rlf-DMRs have been found to overlap regions with active histone marks undertaking chromosome conformation capture would be helpful in determining the target genes of these enhancers. This may in turn identify interactions between genes with Rlf bound at the TSS and Rlf-DMRs.

One caveat to this experiment is that the interactions that may occur within the cell are transient or only present at distinct times in development and not in all tissues/cell types. Also epigenetic modifiers can work together and/ or cross regulate each other differently in different cell types. Thus selecting the correct cell type in addition to developmental time-point is critical for ensuring accurate information is generated. As no previous studies have investigated the importance of Rlf in development and/or disease progression, the identification of a role for Rlf in a specific pathway or disease would greatly influence how future experiments are designed. For example, a putative role for Rlf in cardiac development has been identified in this PhD (refer **Chapter 4**), undertaking chromatin chromosome conformation capture in a cardiac cell type (e.g. cardiomyocytes or epicardial cells), may reveal interactions that occur in cardiac cells that are required for cardiac development. The ability to obtain sufficient amounts of tissue/DNA from early developmental time-points sets limitations on being able to validly perform these experiments.

Chapter 4: Investigating the consequence of loss of Rlf on development

4.1 INTRODUCTION

Mutations in many epigenetic modifiers have been associated with embryonic lethality and/or developmental delay (Daxinger et al., 2013; Okano et al., 1999). Preliminary studies by the Whitelaw laboratory have found *Rlf* mutants to weigh less than their wild-type littermates and are present in reduced numbers at weaning, however no further analysis of phenotype was undertaken (Ashe et al., 2008; Daxinger et al., 2013). Cardiovascular defects are a potential cause of perinatal lethality and several epigenetic modifiers identified in the *Momme* screen have been found to play critical roles in cardiac development (Conway et al., 2003; Hang et al., 2010; Montgomery et al., 2007). The role of *Rlf* in the developing heart will be of particular focus in this chapter.

4.1.1 Epigenetics and heart development

The heart is the first functional organ formed in the developing embryo, and consists of various cell types such as cardiomyocytes, fibroblasts, epicardial and smooth muscle cells (Banerjee et al., 2007). Loss of epigenetic modifiers, such as *Hdac1, Hdac2* and *Lsd1* (lysine specific demethylase 1), have been shown to result in heart defects and perinatal lethality (Montgomery et al., 2007; Nicholson et al., 2013). Global deletion of *Hdac2* results in cardiac defects and perinatal lethality, whilst global deletion of *Hdac1* results in embryonic lethality before E9.5 (Montgomery et al., 2007). Additionally, conditional deletion of both *Hdac1* and

Hdac2 in cardiomyocytes resulted in perinatal lethality due to left ventricle dilation and cardiac stress (Montgomery et al., 2007). Other studies have shown reduced *Lsd1* expression in mice also results in perinatal lethality and ventricular septal defects (Nicholson et al., 2013). Both *Hdac1* and *Hdac2* have been shown to bind the Lsd1-CoREST complex, which plays a role in transcriptional repression (You et al., 2001). Investigations by our lab have also identified members of the Lsd1-CoREST (rest co-repressor 1) complex to be putative binding partners of Rlf *in-vitro* (Harten et al., 2015). Taking these findings into account it is plausible that Rlf may also play a role in cardiac development.

4.1.2 The placenta and cardiovascular development

The placenta also plays an important role in the regulation and development of the cardiovascular system (Thornburg et al., 2010). In some cases abnormal placental development is the underlying cause of abnormal cardiac development and congenital heart disease. For example, mice deficient in *Pparγ* (peroxisome proliferator-activated receptor γ) were found to have placental and cardiac defects (Barak et al., 1999). When *Pparγ* null embryos were aggregated with wild-type placentas the cardiac defect was corrected (Barak et al., 1999). However placental defects are not always the cause of observed cardiac defects. *Pbrm1* (Polybromo 1) is a subunit unique to the Polybromo Associated Factor (PBAF) chromatin remodelling complex, and is expressed in many tissues including the heart and placenta (Wang et al., 2004). The left side of the heart in *Pbrm1* deficient mice was found to be underdeveloped compared to wild-type mice, and an accumulation of trophoblast cells in the placenta was also observed (Wang et al., 2004). It was thought this could

compromise fetal-maternal exchange. When diploid *Pbrm1* null embryos were aggregated with wild-type embryonic stem cells, which contributed to the embryo but not the placenta, the embryonic heart defect was rescued (Wang et al., 2004). These results suggest the defects occurring due to depletion of *Pbrm1* in the heart and placenta are independent of one another, and while a placental defect may be present it may not be the driver of the cardiac defect observed. In this chapter preliminary investigations of the placenta in Rlf mutant mice will also be reported.

This chapter will determine whether *Rlf* is critical for development in the mouse using the Rlf null mouse line *MommeD28*.

4.2 RESULTS

4.2.1 Designing custom antibodies for Rlf

Currently very little is known about Rlf and in which tissues it is expressed. A good quality antibody is important for being able to accurately define in which tissue and/or cell types a protein is expressed, using methods such as Western blotting and immunohistochemistry.

There are two commercial antibodies for Rlf; however each of these has limitations. The first antibody from Atlas (HPA057300) is a polyclonal antibody designed against human RLF that aligns to the middle of the mouse Rlf protein sequence but with poor homology (~63%), **Figure 4.1.** This antibody has been previously tested by other members of the laboratory and has been found to be unsuitable for use in mouse tissue (data not shown). The second commercial antibody from Abcam (AB115011) is also a polyclonal antibody designed against the C-terminal end of human RLF, **Figure 4.1**. The antigen sequence for this antibody also has strong homology to the mouse Rlf protein sequence (~93%).

Western blotting with the Abcam antibody in Rlf tissue has identified multiple bands at both the expected size for Rlf and also at lower molecular weight sizes (data not shown). Additionally, at the beginning of this PhD Abcam began to supply a new lot number for this Rlf antibody which, on testing, was found to be less sensitive and specific than the previously supplied lot number. For these reasons we aimed to design our own mouse Rlf specific antibody to reduce variability and maintain a consistent supply. Three antigen sequences against mouse Rlf were chosen using

GenScript's OptimumAntigen™ design tool to ensure they would be strongly antigenic. The peptide antigens were produced by GenScript, who injected them into rabbits for production of polyclonal antibodies. Antibody containing serum was collected from the rabbits and designated Ab1, Ab2 and Ab3, shown in **Figure 4.1.**

Figure 4.1: Mouse Rlf protein coding sequence.

Alternate exons are identified in blue and black text and the location for each of the Rlf mutations, *MommeD28, MommeD34* and *MommeD8* are indicated by red arrows and green text, and the 16 putative zinc finger domains highlighted in orange. The two commercial antibodies are outlined by a pink text box and the three custom designed antibodies highlighted in black (Ab1), blue (Ab2) and green (Ab3).

4.2.2 Testing unpurified custom designed antibodies for Rlf

Each of the new custom antibodies was first tested unpurified alongside of the Abcam polyclonal Rlf antibody using adult wild-type heart and brain tissue, **Figure 4.2**. Brain tissue was chosen as this has previously been used to validate Rlf expression in wild-type and homozygous fetal mouse tissue (Daxinger et al., 2013). The heart was chosen as an independent tissue that has previously shown high Rlf expression using the Abcam antibody (data not shown).

Initial testing was carried out using adult wild-type tissue due to its ready availability. A single heart and brain tissue sample from one male adult mouse was used for testing of each of the antibodies, to reduce potential variability between samples. The unpurified antibodies were each found to have a higher sensitivity for Rlf (~250 kDa) in the heart than the Abcam antibody which required a longer exposure. No bands were observed in the adult brain samples, suggesting Rlf is not expressed in the adult brain and use of this tissue for validation is not appropriate. Lower molecular weight bands not at the expected size for Rlf were also observed in the heart, ~150 kDa, and brain, ~100 kDa, following blotting with the Abcam antibody, **Figure 4.2.** These lower molecular weight bands were also observed following blotting with each of the unpurified antibodies.

Investigation of annotated Rlf transcripts in the publicly available database Ensembl, Release 83, reveals three potential transcripts for Rlf (Zerbino et al., 2015). These isoforms are predicted to produce Rlf protein of 12, 206, and 217 kDa in size,

suggesting the lower molecular weight bands observed (100 and 150 kDa), may be non-specific bands. Unpurified Ab2 was found to have the most bands present that were not at the expected size for Rlf and it was decided to not continue validation for this antibody, **Figure 4.2**. From these Western blots unpurified Ab1 and Ab3 were chosen for further analysis.

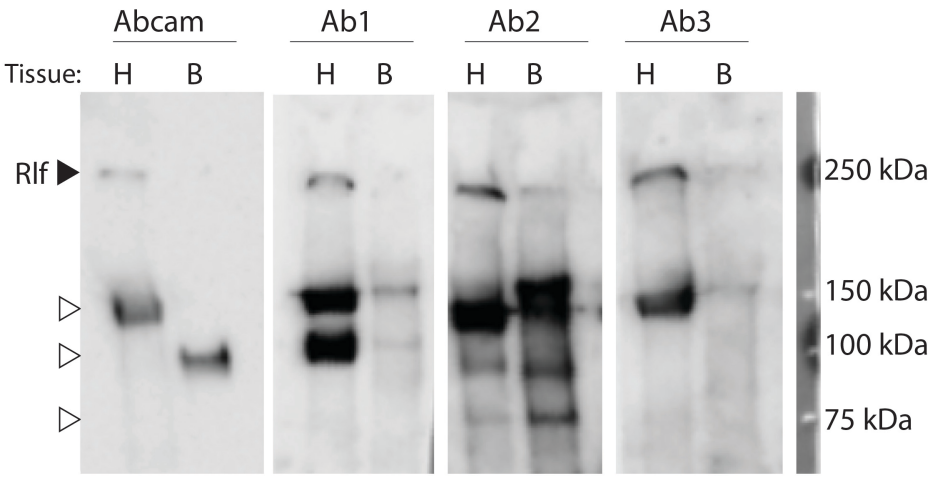


Figure 4.2: Testing of unpurified custom Rlf antibodies.

Western blot of *Rlf* $^{+/+}$ adult heart (H) and brain (B) tissue comparing Abcam polyclonal antibody and unpurified custom antibodies, Ab1, Ab2 and Ab3. Bands at the expected size for Rlf (▶) and not expected size (▷) are indicated.

4.2.3 Testing of purified custom designed antibodies for Rlf

Ab1 and Ab3 were purified according to the GenScript protocol and eluted as either a single fraction (Ab1) or eleven separate fractions (Ab3). The single Ab1 fraction was then tested using *Rlf*$^{+/+}$, *Rlf*$^{MommeD28/MommeD28}$ and *Rlf*$^{MommeD34/MommeD34}$ fetal brain tissue as had been used in the past (Daxinger et al., 2013). One band at the expected size for Rlf was observed in wild-type but not homozygous tissue, **Figure 4.3,** suggesting this is the correct band for Rlf. No bands at other sizes were observed in these samples for Ab1.

Each of the eleven purified fractions for Ab3 were also tested using wild-type fetal head tissue, **Figure 4.4**. Following purification sensitivity of the antibody appeared to be reduced compared to previous Western blots. Of the eleven fractions faint bands at both the expected size for Rlf (~250 kDa) were observed and also bands lower molecular weight than the expected size for Rlf (~100, 90, 75 and 40 kDa each). Fractions 4, 5, 8 and 9 were each found to have a faint a band at the expected size for Rlf, **Figure 4.4**. No bands at the expected size for Rlf or at other sizes were observed in either purified Ab3 fraction 1 or 2, **Figure 4.4**. It was decided to use fractions 4 and 8 for further validation as these fractions were in the middle of the two groups of fractions found to have bands at the expected size for Rlf present.

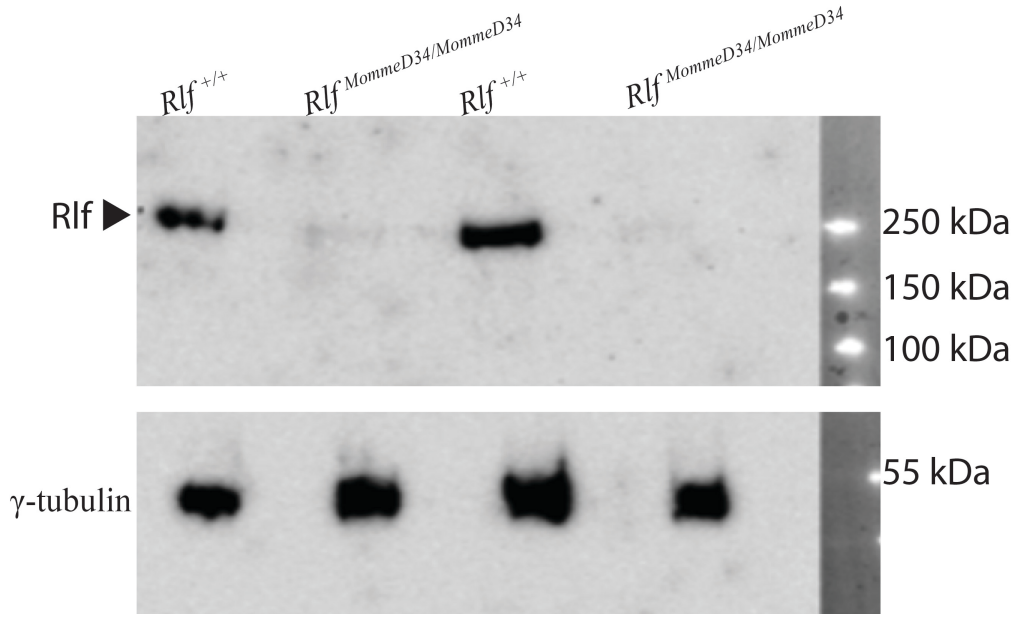


Figure 4.3: Testing of purified custom designed Ab1 for Rlf.

Western blot of purified Ab1 single fraction of *Rlf* $^{+/+}$, *Rlf* $^{MommeD34/MommeD34}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal head showing Rlf protein expression in wild-type but not homozygous tissue.

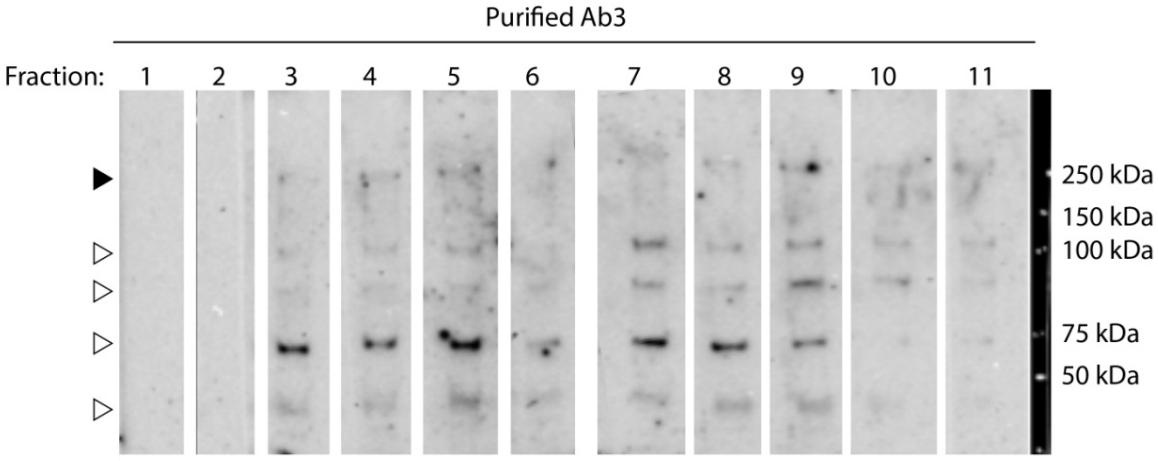


Figure 4.4: Testing of purified custom designed Ab3 for Rlf.

Western blot of 11 fractions purified from Ab3 using *Rlf* $^{+/+}$ fetal head showing Rlf protein expression in wild-type tissue. ▶ Indicates bands at the expected size for Rlf, ▷ indicates bands not at the expected size for Rlf.

Ab3 fractions 4 and 8 (Ab3.4 and Ab3.8, respectively) were tested alongside purified Ab1 and the Abcam antibody to determine which antibody would be the best for use in future experiments. *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal head tissue was used to determine which bands are specific to Rlf. A band at the expected size for Rlf was observed following Western blotting with Ab1, Ab3.4 and the Abcam antibody, but not Ab3.8 in *Rlf* $^{+/+}$ tissue, **Figure 4.5**. In each case this band was not present in *Rlf* $^{MommeD28/MommeD28}$, confirming that these antibodies detect Rlf. The antibodies were also found to detect additional bands at lower molecular weights, which were present in both wild-type and homozygous tissue, **Figure 4.5**, consistent with these bands being non-specific. Of the antibodies tested Ab1 was found to have the greatest sensitivity to Rlf, as such it was decided to use this antibody for future experimentation. However, the presence of multiple bands not at the expected size for Rlf prevents this from being a suitable antibody for immunohistochemistry analysis.

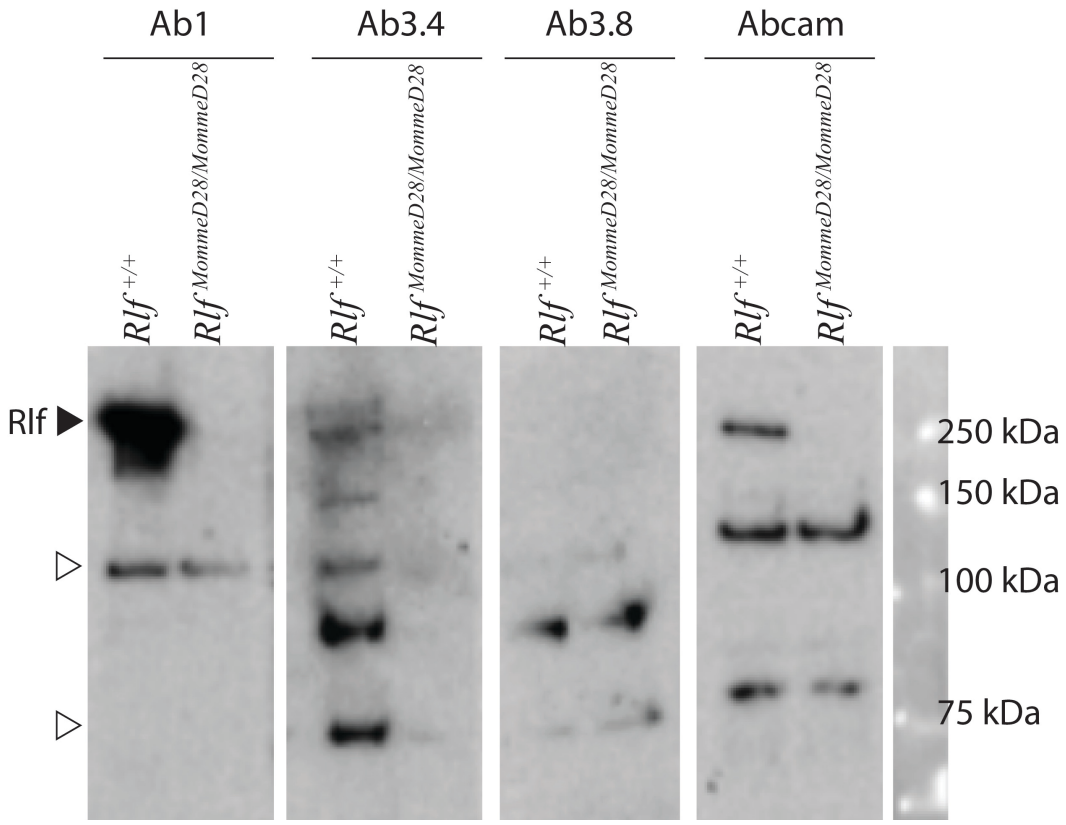


Figure 4.5: Further validation of custom designed antibodies.

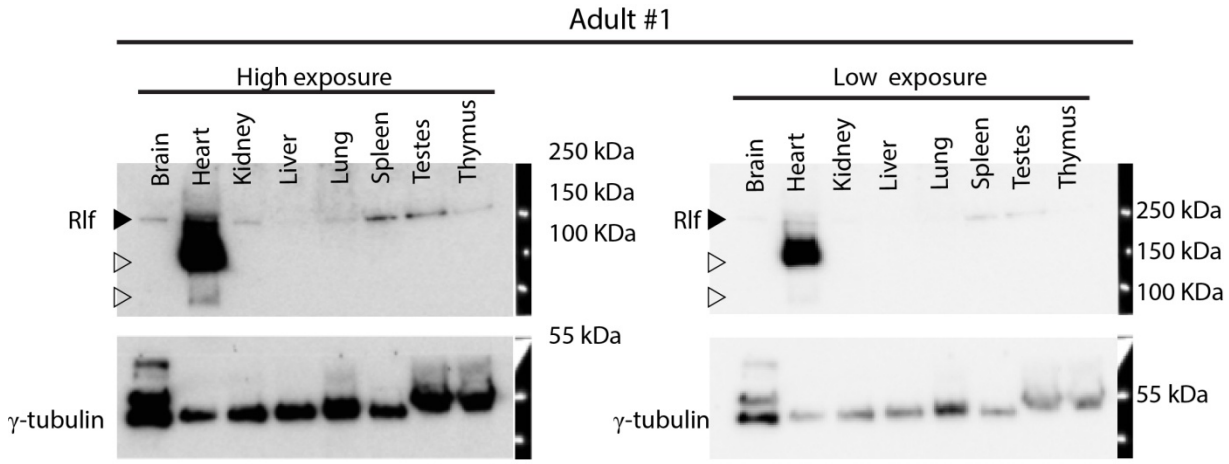
Western blot testing of purified Ab1, purified Ab3 fractions 4 (Ab3.4) and 8 (Ab3.8) and polyclonal Abcam antibody using *MommeD28* wild-type and homozygous fetal head tissue, shows Ab1 has the greatest sensitivity for Rlf and no Rlf expression in mutant tissue. ▶ Indicates bands at the expected size for Rlf, ▷ indicates bands not at the expected size for Rlf.

4.2.4 Rlf is expressed across a range of different tissue

Analysis of human RLF mRNA in adult and fetal tissue by Makela et al., 1995, found RLF is expressed in both adult and fetal heart, brain, spleen, liver, and muscle as well as fetal lung but not adult lung. No studies have analysed the protein expression of human RLF or mouse Rlf. I used the custom designed Rlf antibody, Ab1, for Western blot analysis of a number of fetal, postnatal and adult tissues, as this was found to have a greater sensitivity and specificity than other Rlf antibodies.

Analysis of Rlf expression in four independent eight week old mice found Rlf to be expressed across a range of tissues including brain, heart and liver. Representative blots from two of the mice are shown in **Figure 4.6**. Lower molecular weight bands were also present in heart and muscle tissue (~150 and 100 kDa), but were not observed in any of the other tissues tested. Also of note is weaker bands at the expected size for Rlf were observed in the brain of two mice (**Figure 4.6A**, and data not shown) but not the other two mice studied (**Figure 4.6B**, and data not shown). This suggests Rlf expression in these tissues is near to the lower detection limit of Ab1.

A.



B.

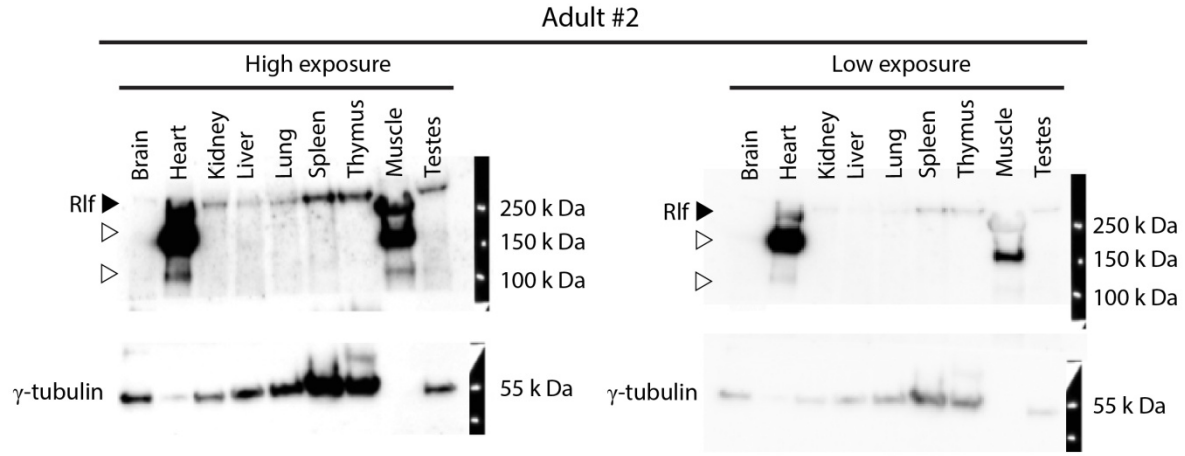


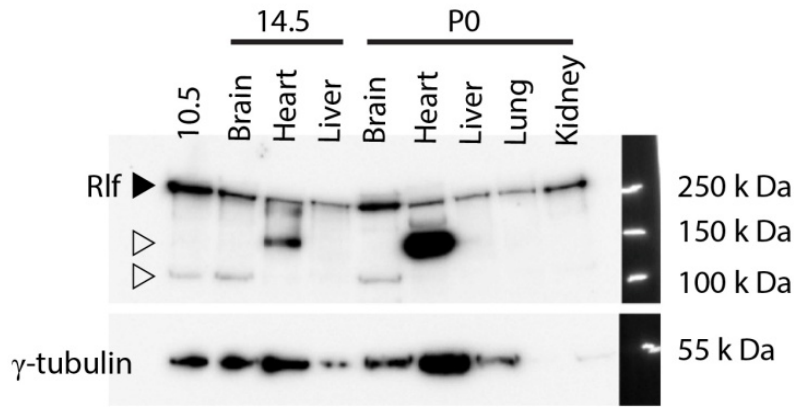
Figure 4.6: Western blotting shows Rlf is widely expressed adult mouse tissue.
Western blot of two independent wild-type adult mice showing Rlf protein expression in different tissues. ▶ Indicates bands at the expected size for Rlf, ▷ indicates bands not at the expected size for Rlf. γ-tubulin was used as a loading control for all Western blots.

Next I asked whether Rlf was expressed in a subset of these tissues at other stages in development. Analysis of whole E10.5 embryos as well as brain, heart, liver (E14.5 and E18.5), and lung and kidney (E18.5 only), also showed Rlf to be widely expressed at these time-points, **Figure 4.7A**.

The lower molecular weight bands previously observed in adult heart tissue were also observed in fetal heart (~150 kDa but not ~100 kDa) and E10.5 whole embryo and fetal brain (~100 kDa each), **Figure 4.7A.** To investigate whether these lower molecular weight bands represented different Rlf isoforms I performed Western blots using $Rlf^{+/+}$, $Rlf^{MommeD28/+}$ and $Rlf^{MommeD28/MommeD28}$ tissue. Postnatal day 0 (P0) brain and heart tissue from the *MommeD28* line were used as positive and negative controls. A reduction in the 250 kDa Rlf band was observed in heterozygous tissue and not present in homozygous tissue as expected, **Figure 4.7B**. No change was observed between the genotypes of the lower molecular weight 150 kDa (heart) and 120 kDa (brain) bands, **Figure 4.7B**.

These results suggest that the 150 kDa and 100 kDa bands are non-specific to Rlf, rather than alternative Rlf isoforms. Taken together these Western blots show that Rlf is widely expressed from an early developmental time-point and is expressed in multiple tissues throughout development and adulthood.

A.



B.

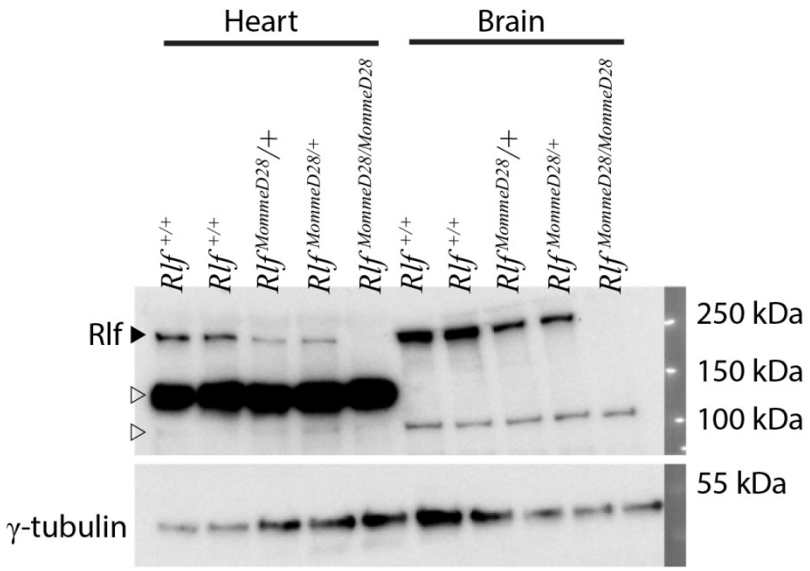


Figure 4.7: Western blotting shows Rlf is widely expressed in fetal mouse tissue.

A. Western blotting shows Rlf protein expression in a number of fetal and postnatal (P0) tissues. B. Western blotting of P0 $Rlf^{+/+}$, $Rlf^{MommeD28/+}$ and $Rlf^{MommeD28/MommeD28}$ heart and brain tissue shows a reduction in expression of Rlf in mutant mice. γ-tubulin was used as a loading control for all Western blots.

4.2.5 *Rlf*$^{MommeD28/MommeD28}$ and *Rlf*$^{MommeD34/MommeD34}$ mice are viable in late gestation

Previous research in the Whitelaw Laboratory found the number of live homozygous *MommeD34* and *MommeD28* mice to be significantly reduced at three weeks (Daxinger et al., 2013). To determine whether these mice were dying before or after birth, I established intercrosses for each of the *Momme* null lines. As dams have been found to give birth at E19.5, embryos were analysed at E18.5 to determine whether *Rlf* mutants are viable at late gestation. Expected Mendelian ratios were observed for both the *MommeD28* and *MommeD34* lines at E18.5 and E14.5, **Table 4.1**. This data, in combination with data collected by other lab members and myself at three weeks (Daxinger et al., 2013), suggests *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ embryos are viable just prior to birth and death is occurring between E18.5 and weaning.

Table 4.1: Offspring from heterozygous intercrosses of *Rlf* mutant mice.

Line	Age	+/+	-/+	-/-	Chi-Test	Reference
Rlf$^{MommeD8/+}$	3 weeks	185 (31)	329 (55)	86 (14)	<0.0001	(Daxinger et al., 2013), this study
Rlf$^{MommeD28/+}$	3 weeks	69 (36)	119 (63)	1 (1)	<0.0001	(Daxinger et al., 2013), this study
Rlf$^{MommeD34/+}$	3 weeks	46 (32)	86 (60)	12 (8)	<0.0001	(Daxinger et al., 2013), this study
Rlf$^{MommeD28/+}$	E18.5	26 (23)	68 (59)	20 (18)	n.s	This study (unpublished)
Rlf$^{MommeD34/+}$	E18.5	18 (31)	28 (48)	12 (21)	n.s	This study (unpublished)
Rlf$^{MommeD28/+}$	E14.5	47 (22)	117 (56)	45 (22)	n.s	This study (unpublished)

Table 4.1: Offspring at 3 weeks, E18.5 and E14.5 from *Rlf Momme* intercrosses. Tabulated data shows the number of observed mice and (% of total), 3 week data (grey) collected by myself and S. Harten (Daxinger et al., 2013), E18.5 and E14.5 (white) data collected by myself.

4.2.6 *Rlf*$^{MommeD28/MommeD28}$ and *Rlf*$^{MommeD34/MommeD34}$ mice weigh less than wild-type littermates

To determine whether there were any gross differences in development between genotypes in each of the Rlf null lines, the weights of E18.5 and E14.5 embryos were measured. Each embryo was carefully patted dry using lint free tissue to remove any excess liquid that may affect measurements.

E18.5 *Rlf* $^{MommeD28/MommeD28}$ embryos were found to weigh significantly less (0.731 g, ± 0.028) than their wild-type (1.144 g, ± 0.019) and heterozygous (1.049 g, ± 0.020) littermates, **Figure 4.8A**. A similar result was also observed in the *MommeD34* line, with *Rlf*$^{MommeD34/MommeD34}$ embryos weighing less (0.767 g, ± 0.035) than both wild-type (1.059 g, ± 0.049) and heterozygous (0.965 g, ± 0.052) littermates, **Figure 4.8B**. The difference in weight at E18.5 between wild-type and homozygous *MommeD28,* **Figure 4.8C**, and *MommeD34* embryos (data not shown) was evident following the removal of embryos from the uterus prior to weighing. These findings are consistent with previously published data at E17.5 (*MommeD34*) and E16.5 (*MommeD28*) (Daxinger et al., 2013). The observation of reduced weight in homozygous and heterozygous offspring from both the *MommeD28* and *MommeD34* mouse lines suggest that alterations in dosage of *Rlf* impact upon embryonic development.

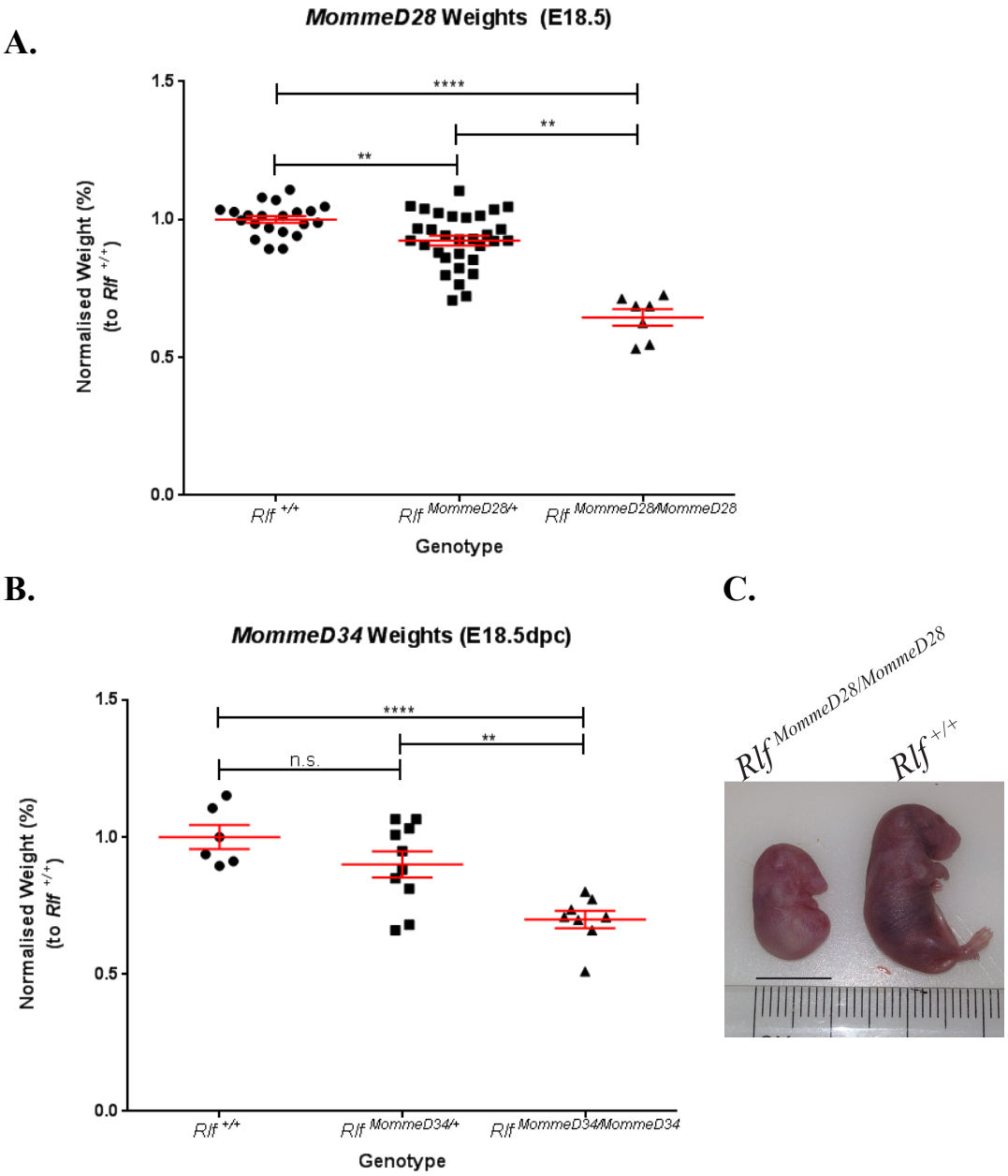
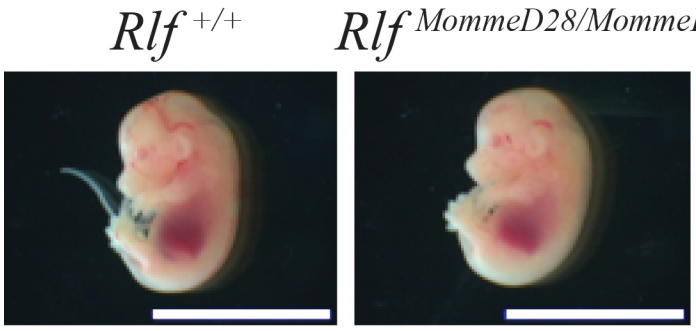


Figure 4.8: E18.5 embryos homozygous for *MommeD28* or *MommeD34* weigh less than wild-type littermates.

A, B. *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD28/+}$(A) and *Rlf* $^{MommeD34/MommeD34}$ and *Rlf* $^{MommeD34/+}$ (B) embryos show a significant decrease in weight versus wild-type littermates at E18.5. n = 7 litters (*MommeD28*), and 3 litters (*MommeD34*). Error bars represent SEM and Student's t-test was used to calculate p values (**** $p <$ 0.0001, ** $p < 0.01$, * $p < 0.05$). Weights for each litter were normalised to the average weight of wild-type littermates, and only litters with more than one wild-type embryo were included. C. Representative *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{+/+}$ E18.5 littermates showing an observable difference in size between genotypes. Scale bar = 1 cm. Each data point represents an individual embryo.

To determine whether the observed difference in weight at E18.5 was present at an earlier developmental time-point *MommeD28* E14.5 embryos were also investigated. Unlike the E18.5 embryos no overt difference in gross appearance between homozygous and wild-type *MommeD28* E14.5 littermates was observed following removal from the uterus, **Figure 4.9A.** *Rlf* $^{MommeD28/MommeD28}$ embryos at this time-point were found to weigh significantly less (0.200 g, ± 0.007) than their wild-type (0.213 g, ± 0.004) and heterozygous (0.222 g, ± 0.003) littermates, **Figure 4.9B**. No difference was observed at this time-point between wild-type and heterozygous littermates, **Figure 4.9B**, suggesting reduced dosage of Rlf does not impact on development until later in gestation. The observation of reduced weight in mid-gestation homozygous embryos provides evidence that *Rlf* plays an important role earlier in embryo development.

A.



B.

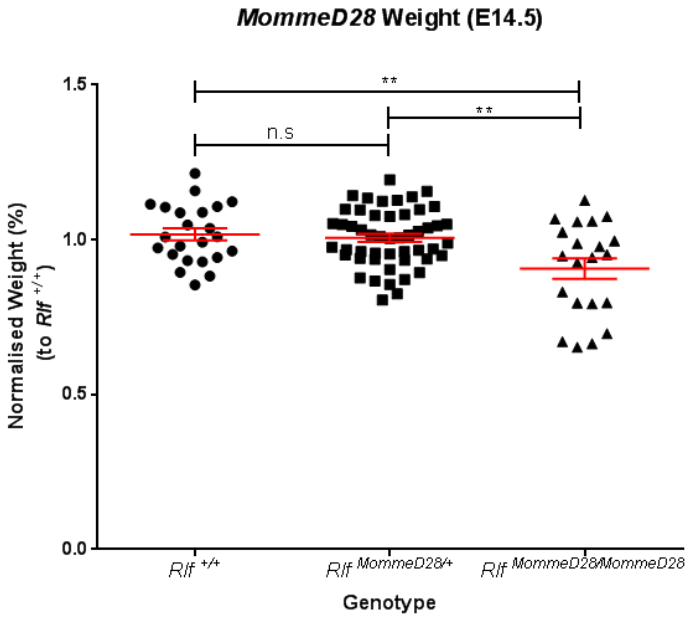


Figure 4.9: E14.5 *Rlf*$^{MommeD28/MommeD28}$ offspring weigh less than their wild-type littermates.

A. No overt difference in size between *Rlf*$^{MommeD28/MommeD28}$ and *Rlf*$^{+/+}$ littermates is observed at E14.5. Scale bar = 5 mm. B. *Rlf*$^{MommeD28/MommeD28}$ embryos at E14.5 from *MommeD28* intercross weigh significantly less than wild-type and heterozygous littermates. Each data point represents an individual embryo. n = 12 litters. Error bars represent SEM and Student's t-test was used to calculate p values (**$p < 0.01$). Weights for each litter were normalised to the average weight of wild-type littermates, and only litters with more than one wild-type embryo were included.

4.2.7 Identification of a heart defect in *Rlf*$^{MommeD28/MommeD28}$ and *Rlf*$^{MommeD34/MommeD34}$ embryos

A number of epigenetic modifiers identified in the *Momme* screen have been found by others to play key roles in cardiac maturation and differentiation and perinatal lethality is often a sign of a cardiovascular defect (Chang & Bruneau, 2012; Conway et al., 2003; Wang et al., 2004). Additionally human infants that are small for gestational age are twice as likely to have congenital heart defects (Malik et al., 2007). As *Rlf* mutants were found to weigh less and have reduced viability (**Figure 4.9; Table 4.1**), I asked whether loss of Rlf in these mice alters cardiac development in E14.5 mid-gestation embryos.

I collected whole E14.5 embryos from the *MommeD28* line for histological analysis (n = 12 per genotype). A tail biopsy was collected for genotyping, heads were removed and bodies fixed in 4% PFA overnight. Embryos were then washed twice with PBS and stored in 70% ethanol. *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ matched littermates were taken to QIMRB histopathology services for embedding in paraffin, before sectioning and undergoing haematoxylin and eosin staining. Whole embryos were used to reduce artefacts that could be introduced from handling of the heart, and to allow for trouble-free orientation, enabling accurate sections at the level of the tricuspid valve to be collected across the cohort. I then analysed the slides blind to genotype to reduce any potential observer bias. Of the 24 embryos analysed all but one were successfully classified into two distinct groups that aligned with their genotype based upon their morphology, **Table 4.2**. One heart, *Rlf* $^{+/+}$10.6, was similar in morphology to its *Rlf* $^{MommeD28/MommeD28}$ littermate and also the other

homozygous mutant hearts analysed. This sample was re-genotyped and confirmed to be a true wild-type embryo, and the difference in morphology in this heart to other wild-types may be independent of the *MommeD28* mutation. Histology images for all 24 hearts analysed are supplied in **Appendix VII**.

Next page:

Table 4.2: E14.5 embryos used to analyse the heart in the *MommeD28* line. All hearts except for one (10.6) separated into two distinct phenotypes that correctly aligned with genotype. Embryo number corresponds to litter number (first) and embryo number within litter (second). V, ventricle; WT, wild-type; M, mutant.

Table 4.2: Observed heart phenotypes in mouse cohort.

Embryo Number	Observations	Predicted Genotype	Genotype
3.1	Thin V. wall and septum with fenestration	M	*Rlf*$^{MommeD28/MommeD28}$
3.3	Well defined V wall and solid septum	WT	*Rlf*$^{+/+}$
4.1	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
4.2	Thin V. walls with deep trabecular recesses	M	*Rlf*$^{MommeD28/MommeD28}$
4.4	Thin V. walls with deep trabecular recesses, also had fenestration at septal base	M	*Rlf*$^{MommeD28/MommeD28}$
4.6	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
5.6	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
5.7	Thin V. walls and septum also thin	M	*Rlf*$^{MommeD28/MommeD28}$
6.7	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
6.8	Thin V. walls with deep trabecular recesses	M	*Rlf*$^{MommeD28/MommeD28}$
7.4	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
7.7	Thin V. wall and V. oddly shaped	M	*Rlf*$^{MommeD28/MommeD28}$
7.9	Thin V. walls with deep trabecular recesses, also had fenestration at septum	M	*Rlf*$^{MommeD28/MommeD28}$
8.6	Thin V. walls with deep trabecular recesses, also had fenestration at septum	M	*Rlf*$^{MommeD28/MommeD28}$
8.7	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
8.8	Thick V. walls and solid septum, some gaps between epicardium and myocardium	WT	*Rlf*$^{+/+}$
10.6	Thin V. wall, whole heart very enlarged	M	*Rlf*$^{+/+}$
10.11	Thin V. walls and fenestrated septum	M	*Rlf*$^{MommeD28/MommeD28}$
12.1	Thin V. wall and thin septum	M	*Rlf*$^{MommeD28/MommeD28}$
12.2	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
14.2	Thick V. walls and solid septum	WT	*Rlf*$^{+/+}$
14.4	Thin V. walls and fenestrated septum	M	*Rlf*$^{MommeD28/MommeD28}$
14.6	Thick V. wall, thin septal base	WT	*Rlf*$^{+/+}$
14.8	Thin V. walls and fenestrated septum	M	*Rlf*$^{MommeD28/MommeD28}$

Histological analysis of *MommeD28* E14.5 embryos identified a potential heart defect in *Rlf*$^{MommeD28/MommeD28}$ embryos, **Figure 4.10** and **Appendix VII**. Slides were analysed in collaboration with Prof R. Harvey, VCCRI, an international expert in cardiac development, and *Rlf*$^{MommeD28/MommeD28}$ hearts were found to display a thin compact layer and over abundance of trabeculae, a subset also displayed fenestrated interventricular septums, **Figure 4.10 i-ii**. These features are similar to ventricular non-compaction phenotypes observed in humans (Prof. R. Harvey, *personal communication*).

As these hearts displayed a thin compact layer I wanted to determine whether there was a difference in ventricle thickness between the two genotypes. Ventricle wall thickness was measured towards the apex of the fetal heart, as is commonly reported in the literature (Vicente-Steijn et al., 2015; Yang et al., 2012). A minimum of five measurements were taken from three serial sections of each embryo and analysed blind to genotype. A significant difference in ventricle wall thickness was observed in both the left and ventricles of *Rlf*$^{MommeD28/MommeD28}$ hearts, supporting the observed differences in morphology, **Figure 4.11A-B.**

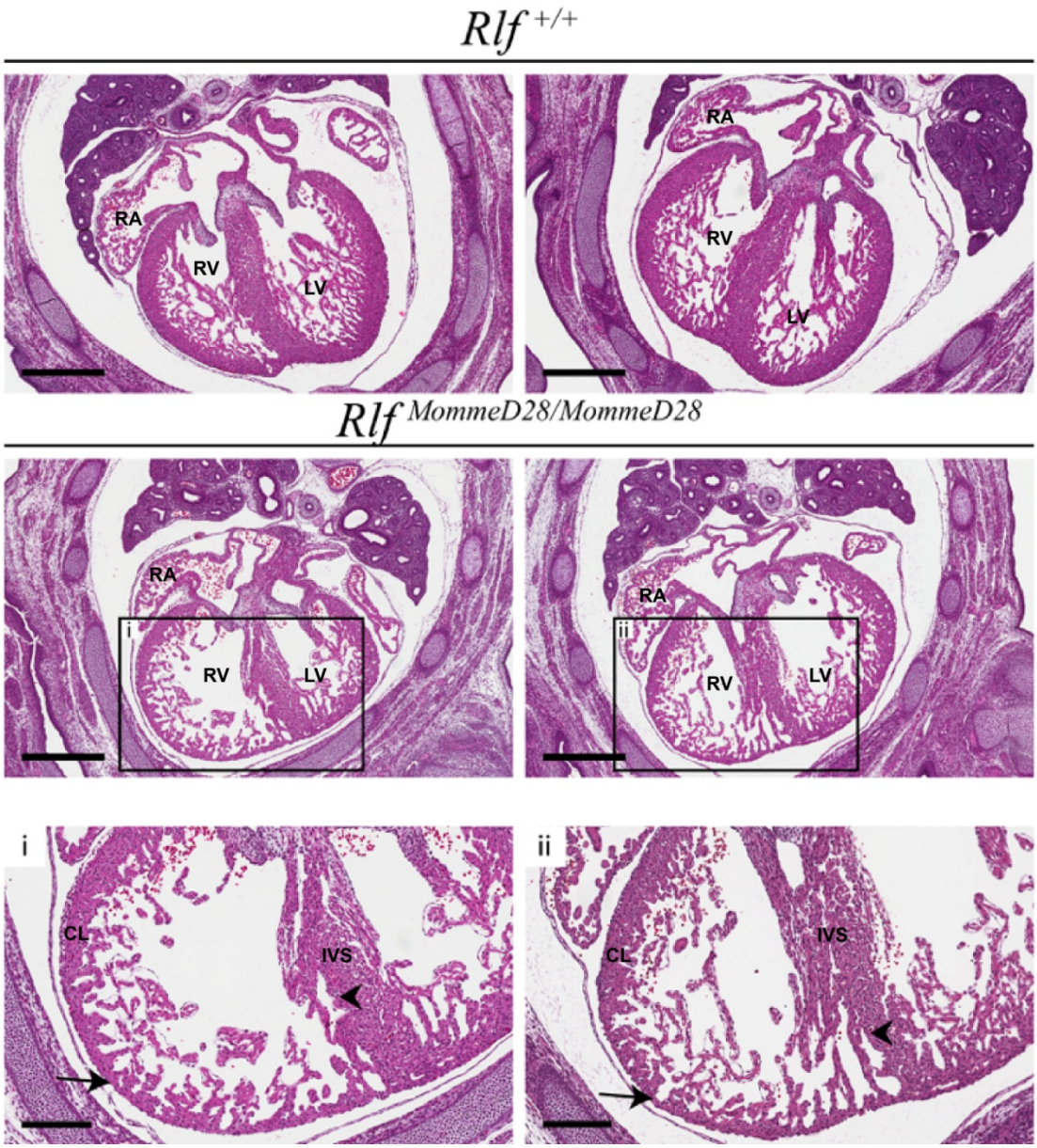


Figure 4.10: *Rlf*$^{MommeD28/MommeD28}$ mutant mice display ventricular and septal defects in the heart at E14.5.

Histological analysis of E14.5 *Rlf* $^{+/+}$and *Rlf* $^{MommeD28/MommeD28}$ embryos. Whole embryos were fixed and sectioned transversely at the tricuspid valve of the heart. i-ii, are high magnification images *Rlf* $^{MommeD28/MommeD28}$ of mutant hearts indicating thin compact layer (arrow), and fenestrated septum (arrowhead). RA, right atria; RV, right ventricle; LV, left ventricle; IVS, interventricular septum; CL, compact layer. *n* = minimum 12 per genotype, scale bar = 500 µm, and 200 µm (i-ii). (Histology of all embryos investigated supplied in Appendix VII).

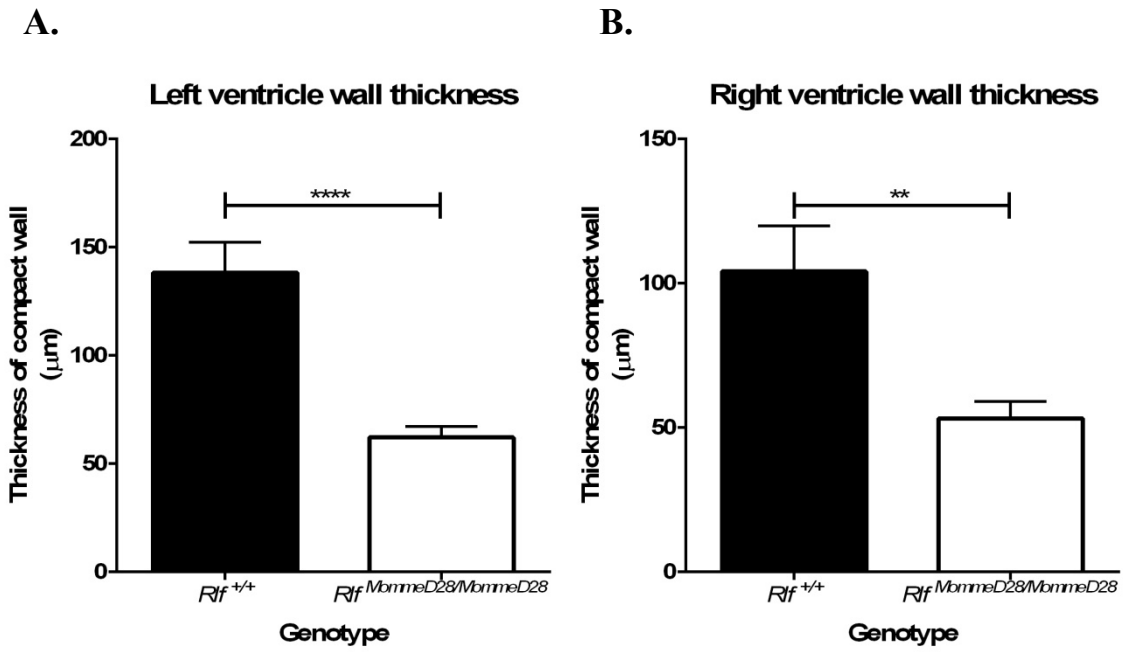


Figure 4.11: *Rlf* $^{MommeD28/MommeD28}$ hearts have thin ventricle walls.

Quantification of ventricle thickness in *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{+/+}$ hearts, showing both the left (A) and right (B) ventricles are thinner in *Rlf* mutants. $n = 12$ per genotype. Error bars represent SEM and Student's t-test was used to calculate p values (**$p < 0.01$, ****$p < 0.0001$).

To confirm whether the defects observed could be attributable to loss of Rlf, I also investigated E14.5 *Rlf* $^{+/+}$ and *Rlf* $^{MommeD34/MommeD34}$ embryos (n = 6 per genotype). Homozygous *MommeD34* embryos were found to display a similar phenotype to that observed in *Rlf* $^{MommeD28/MommeD28}$ embryos, **Figure 4.12** and **Appendix VII**. Again the ventricle walls of *Rlf* $^{MommeD34/MommeD34}$ hearts were found to be significantly thinner than *Rlf* $^{+/+}$ hearts, **Figure 4.13A-B**. The occurrence of the same phenotype in an independent Rlf mutant line suggests that it is loss of Rlf that is underlying the cardiac defect observed. However, it is currently unknown whether this is a heart specific phenotype or the result of developmental delay. Development of a cardiac specific Rlf knockout mouse, or undertaking a time-course analysis to determine whether the defect observed precedes signs of developmental delay, may help to address this in the future.

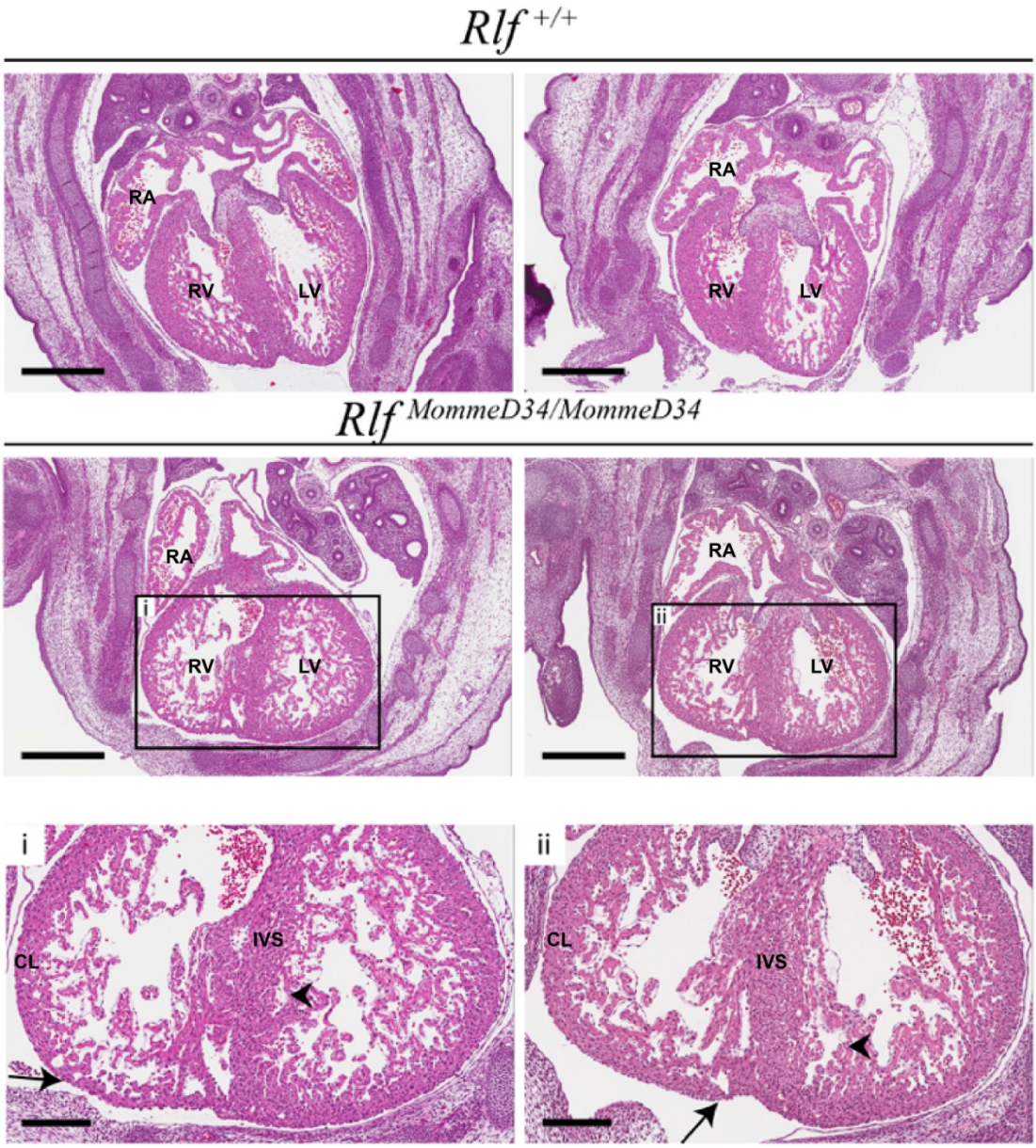


Figure 4.12: *Rlf*$^{MommeD34/MommeD34}$ mutant mice display ventricular and septal defects in the heart at E14.5.

Histological analysis of E14.5 *Rlf* $^{+/+}$and *Rlf* $^{MommeD34/MommeD34}$ embryos. Whole embryos were fixed and sectioned transversely at the tricuspid valve of the heart. i-ii, are high magnification images *Rlf* $^{MommeD34/MommeD34}$ of mutant hearts indicating thin compact layer (arrow), and fenestrated septum (arrowhead). RA, right atria; RV, right ventricle; LV, left ventricle; IVS, interventricular septum; CL, compact layer. *n* = minimum 6 per genotype, scale bar = 500 µm, and 200 µm (i-ii).

(Histology of all embryos investigated supplied in Appendix VII).

A. B.

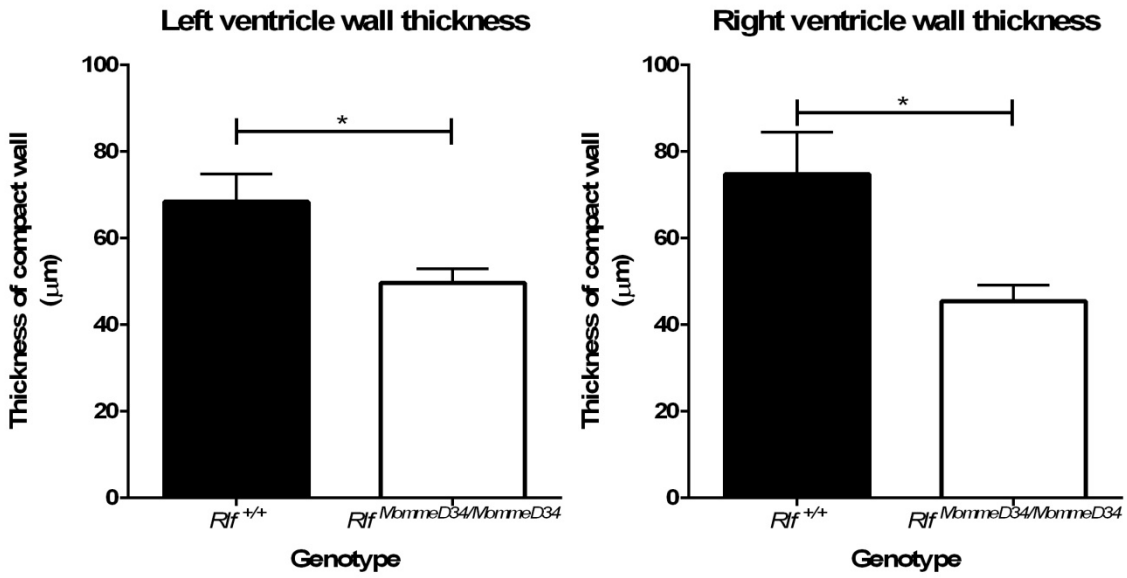


Figure 4.13: *Rlf*$^{MommeD34/MommeD34}$ hearts have thin ventricle walls.

Quantification of ventricle thickness in *Rlf* $^{MommeD34/MommeD34}$ and *Rlf* $^{+/+}$ hearts, showing both the left (A) and right (B) ventricles are thinner in *Rlf* mutants. $n = 6$ per genotype. Error bars represent SEM and Student's t-test was used to calculate p values ($^{*}p < 0.05$).

4.2.8 Analysis of E14.5 placentas reveals no differences in morphology between *MommeD28* genotypes

The placenta has been found to be important in regulating cardiovascular development and fetal growth (Sibley et al., 2005; Thornburg et al., 2010). Additionally, publically available ENCODE data shows Rlf mRNA is expressed in the placenta (data not shown). Following the identification of a potential heart defect I then asked whether loss of Rlf also impacts upon placental morphology. In the mouse the placenta is made up of three layers or zones – the labyrinth zone, junctional zone and maternal decidua, in which the maternal and fetal vasculature is closely intertwined, **Figure 4.14** (Adamson et al., 2002; Rossant & Cross, 2001).

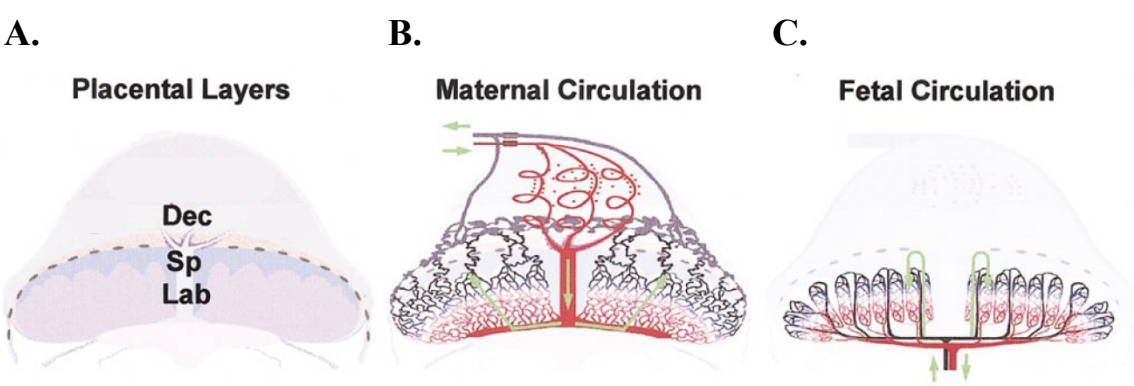


Figure 4.14: Schematic diagrams of zones, maternal and fetal circulation within the murine placenta.

A. Diagram of murine placental layers, indicating the maternal decidua (Dec), spongiotrophoblast layer (Sp) and labyrinth zone (Lab). B, C. Diagram of maternal and fetal circulation respectively, shows the direction of blood flow. Red represents highly oxygenated blood and black least oxygenated blood. Reprinted from Developmental Biology, Vol 250, Issue 2, S.L. Adamson, et al., Interactions between trophoblast cells and the maternal and fetal circulation in the mouse placenta, p. 369, Copyright 2002, with permission from Elsevier. (Appendix XV).

Histology was undertaken on *MommeD28* wild-type and mutant placentas (*n* = min. 5 per genotype) to determine whether there were differences in cellular morphology between the two genotypes. Placentas were fixed in 4% PFA as described previously, and bisected perpendicular to the flat fetal surface prior to embedding and staining. Placental sections were stained with either H&E or Periodic acid Schiff staining (PAS), used to determine the localisation of glycogen containing trophoblast cells, for histology analysis.

Analysis of placentas stained with both H&E and PAS, identified subtle changes in vasculature morphology in placentas from wild-type versus $Rlf^{MommeD28/MommeD28}$ embryos, **Figure 4.15** and **Figure 4.16**. In the mouse maternal vessels can be identified by their trophoblast cell lining and enucleated red blood cells, whilst fetal vessels have a thinner endothelial cell lining and contain nucleated fetal blood cells (Watson & Cross, 2005). No difference in maternal vasculature was observed between the two genotypes, **Figure 4.15** and **Figure 4.16**. However, subtle differences in fetal vasculature were observed, with fetal vessels in $Rlf^{MommeD28/MommeD28}$ derived placentas displaying a more collapsed and disorganised appearance, **Figure 4.15** and **Figure 4.16** (Prof. D. Simmons, *personal communication*). The placentas collected here were sectioned for preliminary phenotyping analysis. However, they were not sectioned appropriately for performing stereology to accurately quantify the phenotype observed. Future work investigating a larger cohort will utilise stereology to accurately quantify these preliminary findings and will be undertaken in collaboration with Prof D. Simmons, University of Queensland, who has expertise in sectioning and fixing placentas for histological analysis.

These preliminary findings in the placenta alongside of the observed reduced weight of *Rlf* mutant embryos, suggest loss of Rlf in the fetal placenta may impact upon embryonic growth due to insufficient circulation. The development of conditional knockout mouse lines would be required to definitively separate the placental defects from the cardiac defects observed this will be discussed further in **Section 4.3.**

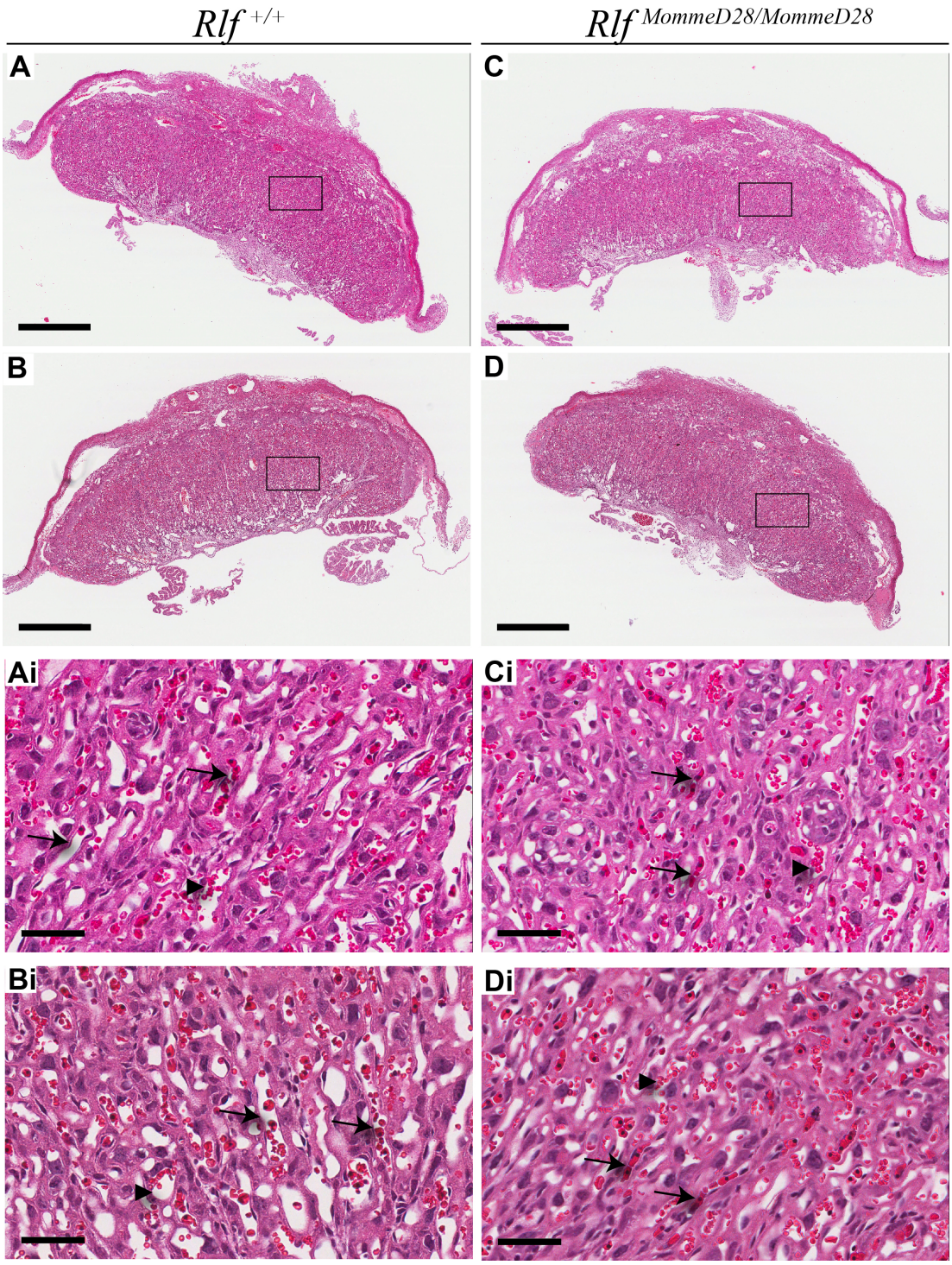


Figure 4.15: H&E analysis of placentas reveals subtle differences in fetal vasculature

Haematoxylin and eosin staining of E14.5 *Rlf* $^{+/}$ and *Rlf* $^{MommeD28/MommeD28}$ placentas. Boxed regions are shown at higher magnification below genotypes (Ai-Di). Fetal vessels, arrows, appear more collapsed and disorganised in *Rlf* $^{MommeD28/MommeD28}$ placentas than wild-type. No difference in maternal vessels, arrowheads, was observed. *n* = min. 5 per genotype. Scale bar = 800 µm (A-D) and 50 µm (Ai-Di).

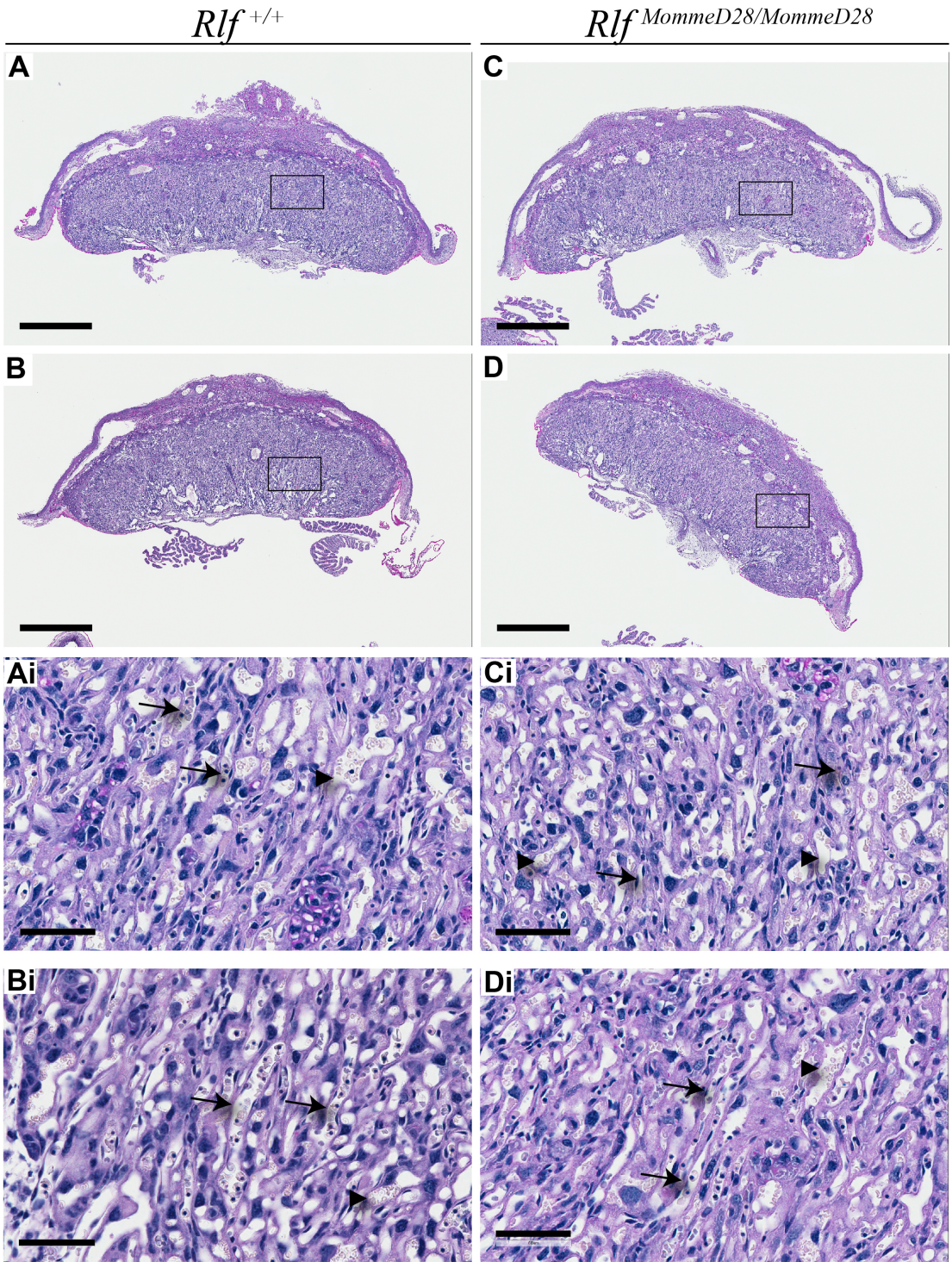


Figure 4.16: PAS staining of placentas reveals subtle differences in fetal vasculature

PAS staining of E14.5 *Rlf* $^{+/}$ and *Rlf* $^{MommeD28/MommeD28}$ placentas. Boxed regions are shown at higher magnification below genotypes (Ai-Di). Fetal vessels, arrows, appear more collapsed and disorganised in *Rlf* $^{MommeD28/MommeD28}$ placentas than wild-type. No difference in maternal vessels, arrowheads, was observed. *n* = min. 5 per genotype. Scale bar = 800 µm (A-D) and 50 µm (Ai-Di).

4.2.9 Subtle differences in cardiac morphology are observed in *Rlf*$^{MommeD28/MommeD28}$ embryos at E11.5

Recent studies from our collaborator, Dr G Del Monte Nieto, VCCRI, have found *Rlf* mRNA is expressed in the epicardium and endocardium as early as E8.0 (Dr G Del Monte Nieto, *personal communication*). Following these findings I asked whether the heart defect observed at E14.5 arises from earlier in development.

I established timed matings of heterozygous *MommeD28* mice to generate embryos at E9.5 and E11.5. As ages of embryos within and between litters can vary dependent on time of implantation, the number of somites was counted for early stage embryos as a measure of developmental age. I removed embryos from the uterus in RNase free cold phosphate buffered saline supplemented with 0.1% Tween (PBT) and counted the number of somites for each embryo. They were then fixed in PFA (made in PBT) overnight at 4 °C, followed by washing in PBT and dehydrated in 50% and 70% ethanol all at 4 °C (all reagents used were RNase free). The yolk sac for each embryo was also collected for genotyping. Following genotyping of embryos, E9.5 and E11.5 *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ embryos were sent to Dr G. Del Monte Nieto, Victor Chang Cardiac Research Institute (VCCRI), who has established technical experience investigating early cardiac development for embedding, sectioning and staining.

Analysis of *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ embryos revealed no difference in morphology at E9.5 – E10.0 (n = min. 13 per genotype) (data not shown, **Appendix VII**). At E11.5 (n = min. 8 per genotype) subtle differences begin to appear in

ventricular morphology between the two genotypes. The left ventricle of *Rlf*$^{MommeD28/MommeD28}$ hearts is more triangular in shape, like a normal right ventricle, making the left and right ventricles appear symmetrical in these hearts, **Figure 4.17** and **Appendix VII.** This finding was consistent amongst homozygous mutant embryos. Another consistent, but subtle, finding in *Rlf*$^{MommeD28/MommeD28}$ hearts is the observation of a less compacted interventricular septum when compared to *Rlf* $^{+/+}$ hearts**, Figure 4.17**. These results suggest the phenotype observed at E14.5 may begin to arise at E11.5 in the developing heart.

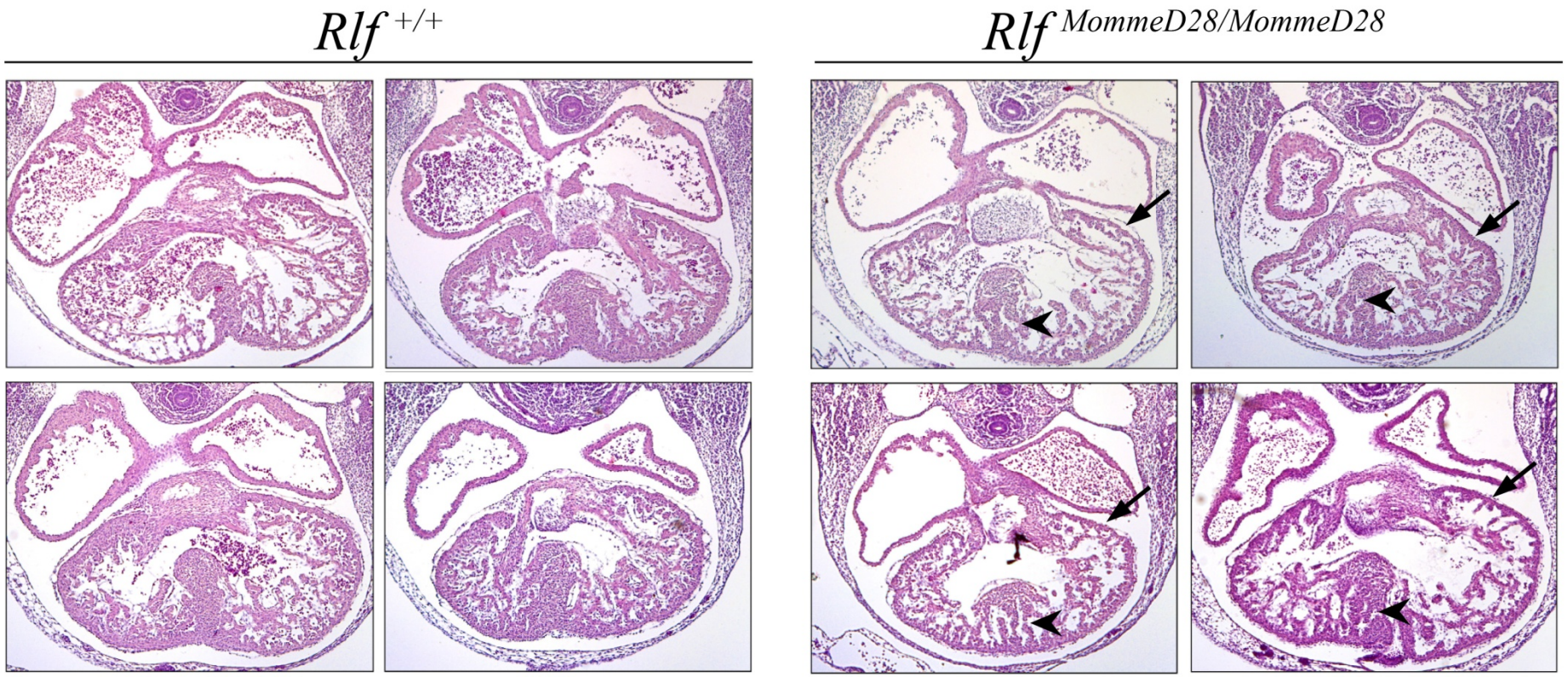


Figure 4.17: Subtle differences in heart morphology are observed as early as E11.5 in *Rlf* mutants.

Rlf $^{MommeD28/MommeD28}$ hearts display less compacted interventricular septum (arrowhead) and symmetrical left ventricle morphology (arrow) (*n* = min. 8 per genotype analysed). (Histology of all embryos analysed supplied in Appendix VII).

4.2.10 Determining the correct samples for differential gene analysis in the heart

After identifying a potential cardiac defect in E14.5 hearts we decided to undertake RNA-seq analysis on whole fetal hearts to identify genes that may be regulated by *Rlf* in the heart. Our intention was to investigate the fetal heart prior to the observation of the heart defect at E14.5, to enable alterations in gene expression that may be causing the changes in morphology to be identified. Preliminary studies in E13.5 embryos found no major defect to be observed at this time-point (data not shown). As such whole E13.5 fetal hearts from the *MommeD28* line were chosen for RNA-seq analysis. However, following the completion RNA-seq analysis in these E13.5 hearts, subtle defects in the heart were observed as early as E11.5 which may impact upon the results reported here. This will be discussed further later in this chapter.

Many papers have addressed the importance of appropriate experimental design when undertaking RNA-seq analysis (Auer & Doerge, 2010; Fang & Cui, 2011; Hansen et al., 2011; Liu et al., 2014). Two considerations when designing this RNA-seq experiment were:

1. Whether to use biological replicates from the same litter regardless of sex, removing the potential of sex specific effects on gene expression being identified

2. To use biological replicates of the same sex, but from different litters

Based on experience from other RNA-seq studies conducted by our laboratory, I decided to use whole E13.5 fetal hearts dissected from littermates. The litter chosen

for RNA-seq consisted of four wild-type, four homozygous and two heterozygous *MommeD28* embryos from the same litter. A third heterozygote from a separate litter was also included and processed alongside of these samples. No gross differences in heart morphology were observed comparing hearts from each genotype, **Figure 4.18.**

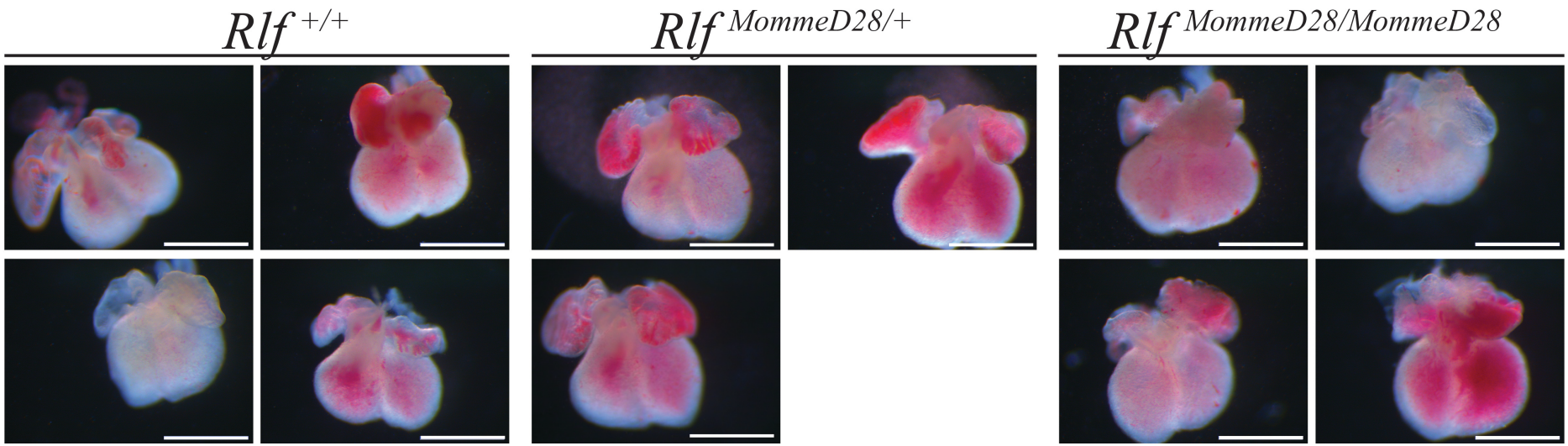


Figure 4.18: No gross differences in heart morphology is observed in *Rlf* mutant hearts at E13.5.

E13.5 whole hearts extracted from *Rlf* $^{+/+}$, *Rlf* $^{MommeD28/+}$ and *Rlf* $^{MommeD28/MommeD28}$ littermates and used for RNA-seq analysis. No overt differences in morphology were observed between genotypes. Embryos were removed from the uterus in cold PBS, and hearts removed from body under dissecting microscope, images were then taken with a Leica DFC320 camera. Scale bar = 1 mm.

4.2.11 RNA extraction and quality validation

I extracted RNA from whole fetal hearts using the AllPrep Micro Kit (Qiagen) as this kit allows for high yields of RNA from small samples. Previous analysis of RNA extracted from whole fetal hearts at this time-point and using this kit found RNA to be of high quality and concentrations to be within the range of 50 – 100 ng/ µl. As a result of this RNA extracted from this cohort was sent directly to the Australian Genome Research Facility (AGRF) for sequencing without calculating the concentration and quality of these samples. As all samples were eluted in a small volume (11 µl) it was decided calculating concentration and quality of these samples by myself and also AGRF could result in excess sample being used leading to reduced DNA being available for sequencing.

Quality control was undertaken by AGRF using the Agilent Bioanalyser RNA 6000 Nano kit, and all samples were found to be of high quality, **Figure 4.19.** The RNA integrity number (RIN) for each sample was > 9.7, where a RIN of 10 is best quality, intact RNA, and a RIN of 1 is totally degraded RNA. The high RIN number for each of these samples makes these appropriate for RNA-seq analysis.

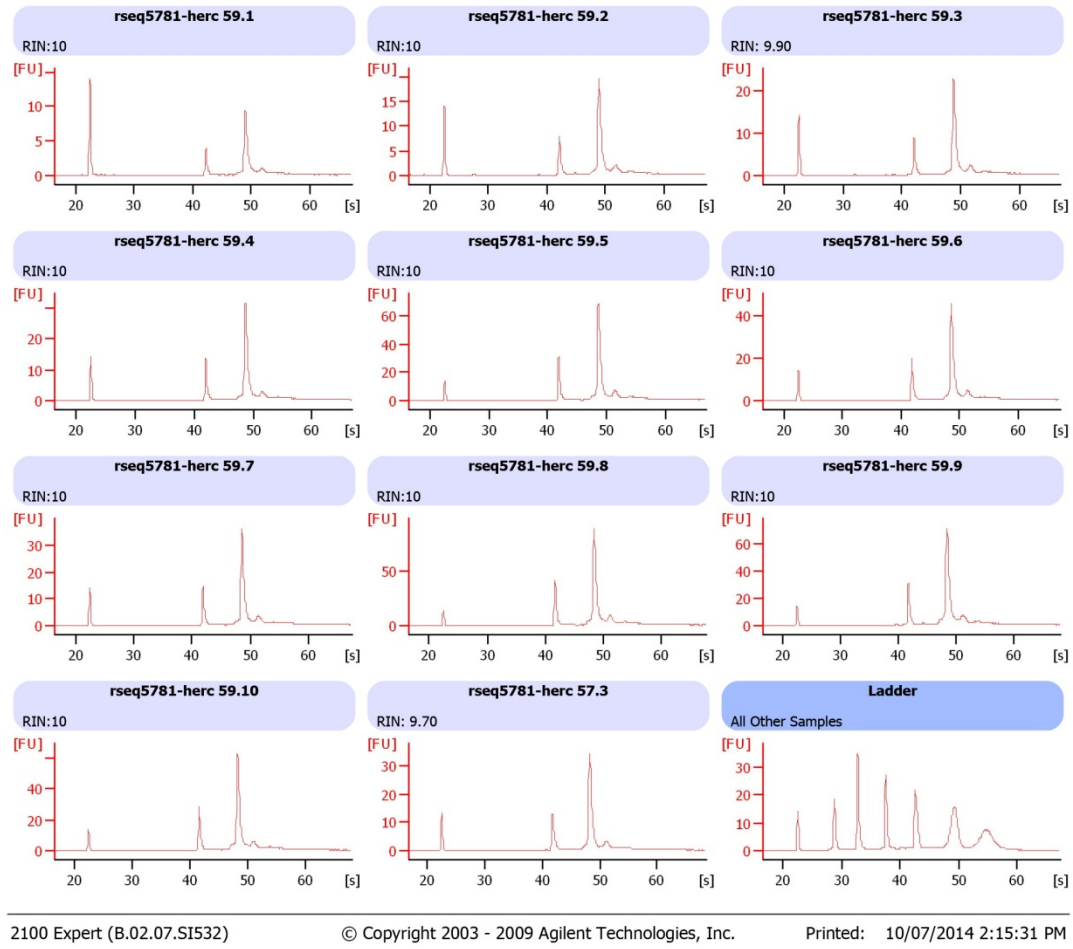


Figure 4.19: Quality of RNA extracted from fetal hearts.

Agilent bioanalyser traces of RNA extracted from whole fetal hearts showing each contains high quality RNA for sequencing.

4.2.12 RNA sequencing and mapping

mRNA library preparation and Illumina HiSeq 2000 sequencing, on polyadenylated RNA, with 50 bp single end reads was performed by the AGRF. Each sample was run on two lanes for sequencing and all were found to meet AGRF quality standards, the data yield for each sample and lane is detailed in **Table 4.3**. Following sequencing I undertook quality control checks on this raw data using FastQC (Andrews, 2010). FastQC reports on the following measurements: per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences and Kmer content. All sequencing files successfully passed FastQC analysis.

Table 4.3: RNA sequencing data yield.

Lane	Sample Name	Genotype	Single Reads	Data Yield (bp)
6	Herc_59-1_RNA	*Rlf*$^{+/+}$	14,873,339	0.74 Gb
	Herc_59-2_RNA	*Rlf*$^{+/+}$	17,398,715	0.87 Gb
	Herc_59-3_RNA	*Rlf*$^{MommeD28/MommeD28}$	17,307,986	0.87 Gb
	Herc_59-4_RNA	*Rlf*$^{MommeD28/MommeD28}$	16,505,467	0.83 Gb
	Herc_59-5_RNA	*Rlf*$^{+/+}$	18,269,620	0.91 Gb
	Herc_59-6_RNA	*Rlf*$^{MommeD28/MommeD28}$	17,831,375	0.89 Gb
	Herc_59-7_RNA	*Rlf*$^{+/+}$	18,148,076	0.91 Gb
	Herc_59-8_RNA	*Rlf*$^{MommeD28/MommeD28}$	18,175,123	0.91 Gb
	Herc_59-9_RNA	*Rlf*$^{MommeD28/+}$	18,182,655	0.91 Gb
	Herc_59-10_RNA	*Rlf*$^{MommeD28/+}$	15,840,263	0.79 Gb
	Herc_57-3_RNA	*Rlf*$^{MommeD28/+}$	15,052,975	0.75 Gb
7	Herc_59-1_RNA	*Rlf*$^{+/+}$	14,861,472	0.74 Gb
	Herc_59-2_RNA	*Rlf*$^{+/+}$	17,410,231	0.87 Gb
	Herc_59-3_RNA	*Rlf*$^{MommeD28/MommeD28}$	17,251,411	0.86 Gb
	Herc_59-4_RNA	*Rlf*$^{MommeD28/MommeD28}$	16,476,072	0.82 Gb
	Herc_59-5_RNA	*Rlf*$^{+/+}$	18,231,528	0.91 Gb
	Herc_59-6_RNA	*Rlf*$^{MommeD28/MommeD28}$	17,840,737	0.89 Gb
	Herc_59-7_RNA	*Rlf*$^{+/+}$	18,141,441	0.91 Gb
	Herc_59-8_RNA	*Rlf*$^{MommeD28/MommeD28}$	18,070,558	0.90 Gb
	Herc_59-9_RNA	*Rlf*$^{MommeD28/+}$	18,200,960	0.91 Gb
	Herc_59-10_RNA	*Rlf*$^{MommeD28/+}$	15,824,463	0.79 Gb
	Herc_57-3_RNA	*Rlf*$^{MommeD28/+}$	15,048,856	0.75 Gb
Total			**374,943,323**	**18.73 Gb**

Table 4.3: Data sequencing yield for each replicate in two lanes showing no bias in reads towards any particular replicate.

All of the following bioinformatics analysis was undertaken by myself using the open-web based platform Galaxy (www.galaxy-qld.genome.edu.au) (Blankenberg et al., 2001; Giardine et al., 2005; Goecks et al., 2010). Reads were aligned using Tophat (Trapnell et al., 2009), a program that aligns RNA-seq reads to the genome using the short read aligner BowTie, and analyses the mapped results to identify splice junctions. The following Tophat parameters were used: maximum intron length between donor/acceptor site must be less than 100,000 bp; coverage search disabled; read alignments with > 2 mismatches discarded; library type is unstranded

Read counts for mRNA transcripts were then extracted using HTSeq count matrix (Anders & Huber, 2010). This tabulates mapped files into a matrix form for use in differential gene expression tools. The following HTSeq count parameters were used: not from a strand specific assay; model for counting reads over the supplied gene model set to intersection strict; gene annotations used from Ensembl (www.ensembl.org, release 67).

4.2.13 Differential gene expression analysis

Currently, there are many different methods available for studying differential gene expression, with each method differing in their normalisation methods, isoforms expression analysis and correction of multiple testing (Rapaport et al., 2013). I chose to analyse differential gene expression in fetal hearts using three common methods: CuffDiff (Trapnell et al., 2012), DESeq (Anders & Huber, 2010) and EdgeR (Robinson et al., 2010), each of which utilise the Benjamini-Hochberg method for controlling false discovery rate (Benjamini & Hochberg, 1995). Default parameters were used for each program. Genes were considered differentially expressed if they had an adjusted p value of < 0.005 (after false discovery correction) and a fold change > 1.2. CuffDiff was found to be the most stringent with 228 genes identified as differentially expressed in *Rlf* $^{MommeD28/MommeD28}$ hearts (**Appendix VIII**), followed by EdgeR with 293 genes (**Appendix IX**) and DEseq 316 genes (**Appendix X**). **Figure 4.20** shows the agreement between the three analysis methods, with 162 genes identified as differentially expressed in all three programs. In each of the programs slightly more genes were found to be down-regulated versus up-regulated in *Rlf* $^{MommeD28/MommeD28}$ fetal hearts, **Table 4.4**, correlating with previous data suggesting *Rlf* is required for transcriptional activation.

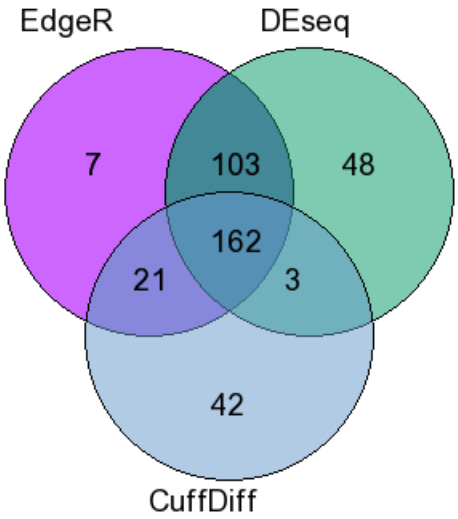


Figure 4.20: Agreement of differential gene expression across methods.
Venn diagram showing intersection of genes identified as differentially expressed in three different programs – CuffDiff, EdgeR and DEseq.

Table 4.4: **More genes are down-regulated in *Rlf*$^{MommeD28/MommeD28}$ hearts than up-regulated**

		Up-regulated	Down-regulated
CuffDiff ($p < 0.005$)	> 5 FC	1	2
	> 2 FC	8	9
	> 1.5 FC	42	55
	> 1.2 FC	98	130
DEseq ($p < 0.005$)	> 5 FC	0	0
	> 2 FC	0	7
	> 1.5 FC	31	66
	> 1.2 FC	111	205
EdgeR ($p < 0.005$)	> 5 FC	1	3
	> 2 FC	11	38
	> 1.5 FC	57	99
	> 1.2 FC	108	186

Table 4.4: Each program used to analyse differential gene expression revealed more genes to be down-regulated in homozygous mutant hearts. All *p* values were corrected for multiple testing (Full list of genes Appendices VIII, IX, X).

4.2.14 Clustering of biological replicates

EdgeR also allows for the similarity of each biological replicate to be analysed using multi dimensional scaling. Three of the wild-types (i - iii) were found to cluster tightly together; however one wild-type embryo (iv) was separated from the group, **Figure 4.21A**. A likely reason for this may be due to a slight difference in overall body morphology observed between this embryo and its wild-type littermates, **Figure 4.21B**. Removal of this sample and repeating the EdgeR bioinformatic analysis with three wild-type replicates found no impact upon the list of differentially expressed genes. *Rlf* $^{MommeD28/MommeD28}$ hearts were found to cluster into two groups, independent of sex, (a and b; c and d), **Figure 4.21A-B**. Again removal of individual replicates had no impact upon gene expression analysis.

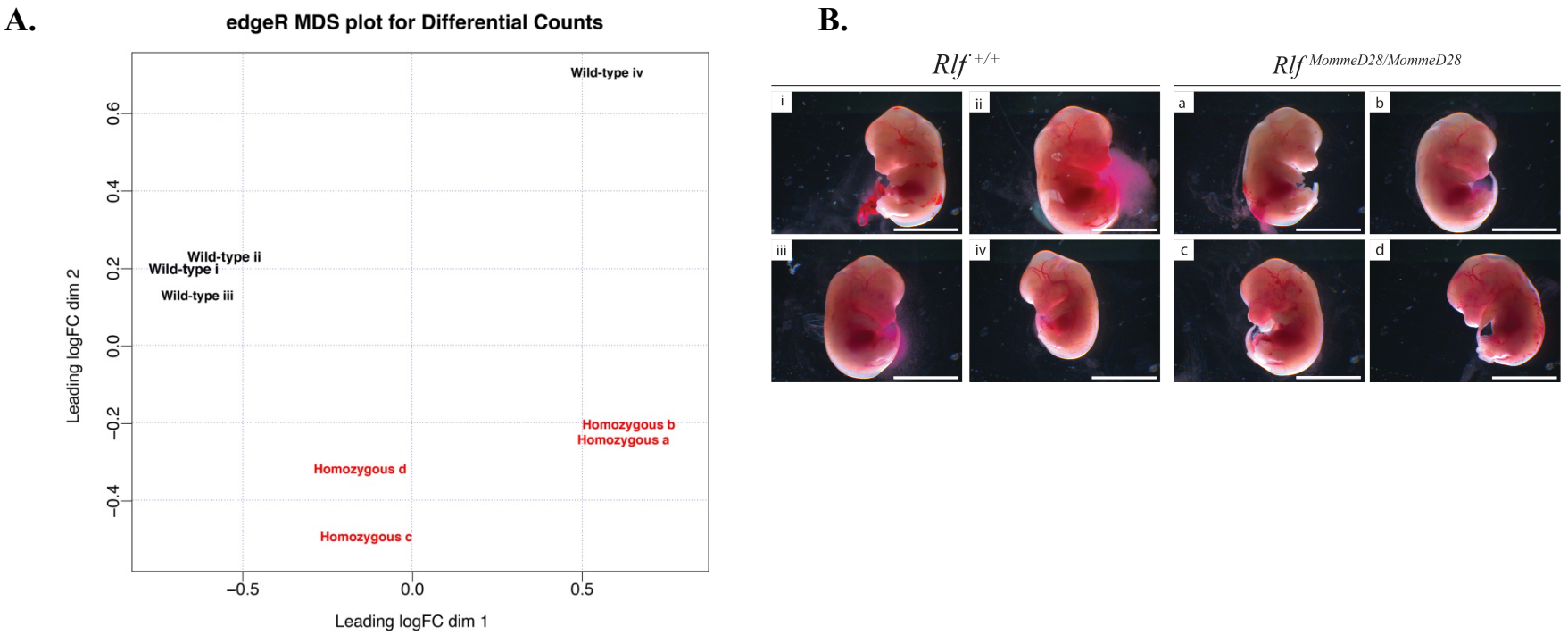


Figure 4.21: Replicate clustering and similarity of *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ samples.

A. EdgeR multidimensional scaling (MDS) plot displaying the similarity amongst samples, following analysis. The four homozygous mutants cluster into two groups of two (red), whilst one wild-type sample did not cluster near the other three wild-type samples (black). B. Gross morphology of wild-type (i-iv) and homozygous (a-d) embryos use for RNA-seq analysis. i, ii and iii clustered together, but not iv; homozygous embryos a, b clustered together whilst c and d clustered separately. Embryos were removed from the uterus in cold PBS images taken with a Leica DFC320 camera. Scale bar = 5 mm.

4.2.15 Differentially expressed genes in the fetal heart

Table 4.5 and **Table 4.6**, show the top 20 significantly up- and down-regulated genes between *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{+/+}$ E13.5 fetal hearts as calculated by EdgeR using a *p* value < 0.005 and fold change > 1.5 (for full list refer **Appendix IX**). Analysis of up- and down-regulated genes found them to be associated with several pathways including Notch signalling and cardiac hypertrophy. Notch pathway genes found to be altered by loss of *Rlf* included those encoding Jagged1 (*Jag1*), Jagged2 (*Jag2*), *Hey1*, *Cntn1* and *Dll1* (**Figure 4.22A**). Ingenuity's Upstream Regulator Analysis identified Notch1 (z score = -2.356), Myod1 (z = n-2.132), and Csll1 (z = -2.00) as inhibited upstream regulators, and Gata1 (z = 2.393), Rnf2 (z = 2.236), and Nrli2 (z = 2.219) as activated upstream regulators (**Figure 4.22B**). *Vegfa* and Conexin40 (*Gja5*), genes important for heart chamber development, were also found to be differentially expressed.

Pages 195 and 196:

Table 4.5: Top 20 down-regulated genes respectively in Rlf $^{MommeD28/MommeD28}$ fetal heart compared to wild-type following RNA-seq analysis. Fetal heart was collected from four individual E13.5 embryos for each genotype; all embryos were from the same litter. RNA-seq was performed by AGRF on the Illumina HiSeq 2000 and bioinformatic analysis in EdgeR by myself. Adjusted p value controls for false discovery rate through the use of the Benjamini-Hochberg procedure. (Full list **Appendix IX**).

Table 4.6: Top 20 up-regulated genes respectively in *Rlf* $^{MommeD28/MommeD28}$ fetal heart compared to wild-type following RNA-seq analysis. Fetal heart was collected from four individual E13.5 embryos for each genotype; all embryos were from the same litter. RNA-seq was performed by AGRF on the Illumina HiSeq 2000 and

bioinformatic analysis in EdgeR by myself. Adjusted p value controls for false discovery rate through the use of the Benjamini-Hochberg procedure (Full list **Appendix IX**).

Table 4.5: Top 20 Down-regulated Genes in *Rlf*$^{MommeD28/MommeD28}$ Fetal Heart.

Symbol	Gene Name	Log Fold Change	Adjusted P value
Apoc1	Apolipoprotein C-I	18.28	4.50E-18
Tex11	Testis expressed gene 11	10.64	8.49E-12
Cpa1	Carboxypeptidase A1	9.64	3.56E-31
Tdrd1	Tudor domain containing 1	4.58	2.29E-06
Ap3b2	Adaptor-related protein complex 3, beta 2 subunit	4.28	5.95E-35
Cpa4	Carboxypeptidase A4	3.95	6.83E-08
Slc36a2	Solute carrier family 36 (proton/amino acid symporter), member 2	3.72	1.29E-10
Cilp2	Cartilage intermediate layer protein 2	3.47	1.83E-06
Pcdh10	Protocadherin 10	3.24	2.03E-08
Atp2a1	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	3.22	7.84E-05
Col2a1	Collagen, type II, alpha 1	3.13	1.85E-43
Zfp783	Zinc finger protein 783	2.97	5.75E-07
Lrrtm4	Leucine rich repeat transmembrane neuronal 4	2.90	4.04E-07
Gm1564	Predicted gene 1564	2.89	2.86E-06
Tesc	Tescalcin; similar to Tescalcin	2.86	5.72E-07
Pcdhb7	Protocadherin beta 7	2.80	1.87E-10
Tnnt3	Troponin T3, skeletal, fast	2.78	2.18E-06
Pilra	Paired immunoglobin-like type 2 receptor alpha	2.65	1.85E-06
Mal	Myelin and lymphocyte protein, T-cell differentiation protein	2.59	1.89E-05
Acta1	Alpha 1, skeletal muscle	2.58	3.14E-33

Table 4.6: Top 20 Up-regulated Genes in *Rlf*$^{MommeD28/MommeD28}$ Fetal Heart.

Symbol	Gene Name	Log Fold Change	Adjusted P value
Prap1	Proline-rich acidic protein 1	9.22	5.63E-09
Aldh1a7	Aldehyde dehydrogenase family 1, subfamily A7	3.08	1.82E-09
Treml1	Triggering receptor expressed on myeloid cells-like 1	2.60	4.40E-06
Gpat2	RIKEN cDNA A530057A03 gene	2.59	0.000101
Vtcn1	V-set domain containing T cell activation inhibitor 1	2.46	2.62E-05
Ppbp	Pro-platelet basic protein	2.40	9.21E-07
Gdf10	Growth differentiation factor 10	2.38	0.001085
Tubb1	Tubulin, beta 1	2.32	1.19E-05
Gp9	Glycoprotein 9	2.29	3.26E-05
Ptprq	Protein tyrosine phosphatase, receptor type, Q	2.26	1.73E-05
Gp1bb	Glycoprotein Ib, beta polypeptide	2.19	3.37E-05
Gdpd3	Glycerophosphodiester phosphodiesterase domain containing 3	2.14	0.000721
Gstm6	Glutathione S-transferase	2.10	0.000118
Pf4	Platelet factor 4	2.09	2.18E-10
Ptgds	Prostaglandin D2 synthase (brain)	2.07	0.000136
Il23a	Interleukin 23, alpha subunit p19	1.99	2.47E-05
Fcgr2b	Fc receptor, IgG, low affinity IIb	1.98	0.000148
Hpdl	4-hydroxyphenylpyruvate dioxygenase-like	1.97	0.001163
Rinl	Ras and Rab interactor-like	1.96	2.73E-06
Cd14	CD14 antigen	1.96	9.47E-05

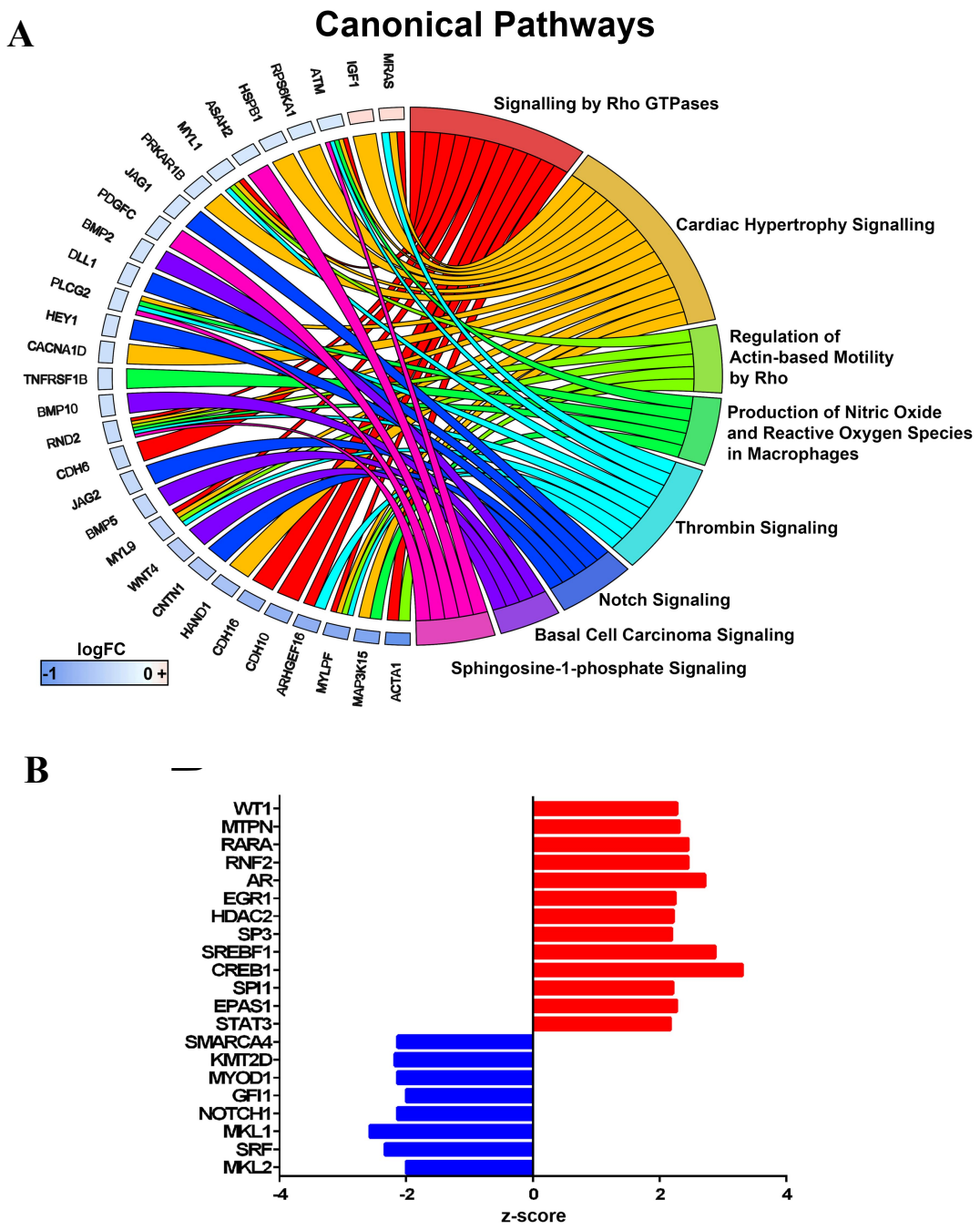


Figure 4.22: Transcriptome analysis shows dysregulation of the NOTCH signalling pathway in *Rlf*$^{MommeD28/MommeD28}$ hearts

A. GOplot representation of the analysis of canonical pathway enriched in the 511 differentially expressed genes in *Rlf*$^{MommeD28/MommeD28}$ fetal hearts. B-H P value, Benjamini-Hochberg-adjusted P value. **B.** Putative upstream regulator analysis showing activated upstream regulators (red) and inhibited upstream regulators (blue), predicted utilising differentially expressed genes identified following RNA-seq analysis (511 genes).

Recent studies undertaken by our collaborator, investigating the expression of a subset of Notch target genes identified in RNA-seq, confirmed their dysregulation following *in-situ* hybridisation analysis (Dr G Del Monte Nieto, *personal communication*). Additionally, immunohistofluorescence of activated Notch1 found no expression of activated Notch1 in *Rlf* $^{MommeD28/MommeD28}$ mutant hearts, whilst expression in wild-type hearts was normal (Dr G Del Monte Nieto, *personal communication*). These findings support the RNA-seq and pathway analysis performed suggesting *Rlf* is an important requirement for the developing heart.

Furthermore, *in-situ* hybridisation analysis for *Rlf* mRNA expression has also found *Rlf* to be expressed only in the endocardium and epicardium of the heart (Dr G Del Monte Nieto, *personal communication*). These two cell types make up a small percentage of the total cell population within the heart (~15%) (Banerjee et al., 2007). This may account for the small fold-changes observed in Notch1 target genes in our data being a result of dilution, as multiple cardiac cell types were analysed.

Ideally, undertaking RNA-seq in a single cardiac cell type, such as endocardial cells in which *Rlf* expression has been observed, may be more informative. This may reveal specific changes in gene expression earlier in development that lead to the heart defect observed in *Rlf* $^{MommeD28/MommeD28}$ hearts. However, performing RNA-seq in a single cardiac cell type from the developing heart would be technically difficult due to the small amount of tissue/RNA available. Recently, commercial kits have been designed that allow for immunomagnetic isolation of cardiomyocytes and have been successfully used in research applications (O'Meara et al., 2015). However, a

significant number of hearts (5 – 10) are still required to generate one replicate that is suitable for RNA-seq analysis. The further development of such kits for isolating epicardial and endocardial cells, and improvements in single cell RNA-seq would greatly enhance our understanding of how loss of *Rlf* influences the cardiac phenotype observed.

4.2.16 Investigation of a fetal liver Rlf-DMR that overlaps an active enhancer mark in the heart

As we had previously noted our fetal liver Rlf-DMRs overlapped regulatory regions that are active in other tissue types, I asked whether any overlapped with regulatory regions in the fetal heart. Almost half of the Rlf-DMRs were found to overlap with at least one histone mark associated with active and or poised regulatory elements (589 of 1,329 Rlf-DMRs), **Figure 4.23A**. Some of which have been validated as active enhancers in transgenic mouse assays (data not shown) (Visel et al., 2007). I chose to study one of the candidate Rlf-DMRs that was found to overlap with the active enhancer marks H3K4me1 and H3K27ac in the fetal heart, **Figure 4.23B**. This Rlf-DMR overlapped a previously identified active enhancer, mm87, located on Chromosome 10: 44,948,419 - 44,951,592 (Visel et al., 2007). This enhancer, showed cardiac specific expression at E11.5 in seven embryos, a representative embryo is shown in **Figure 4.23C** (Visel et al., 2007). Making this a suitable candidate for determining whether the Rlf-DMR observed in the fetal liver is also present in the fetal heart.

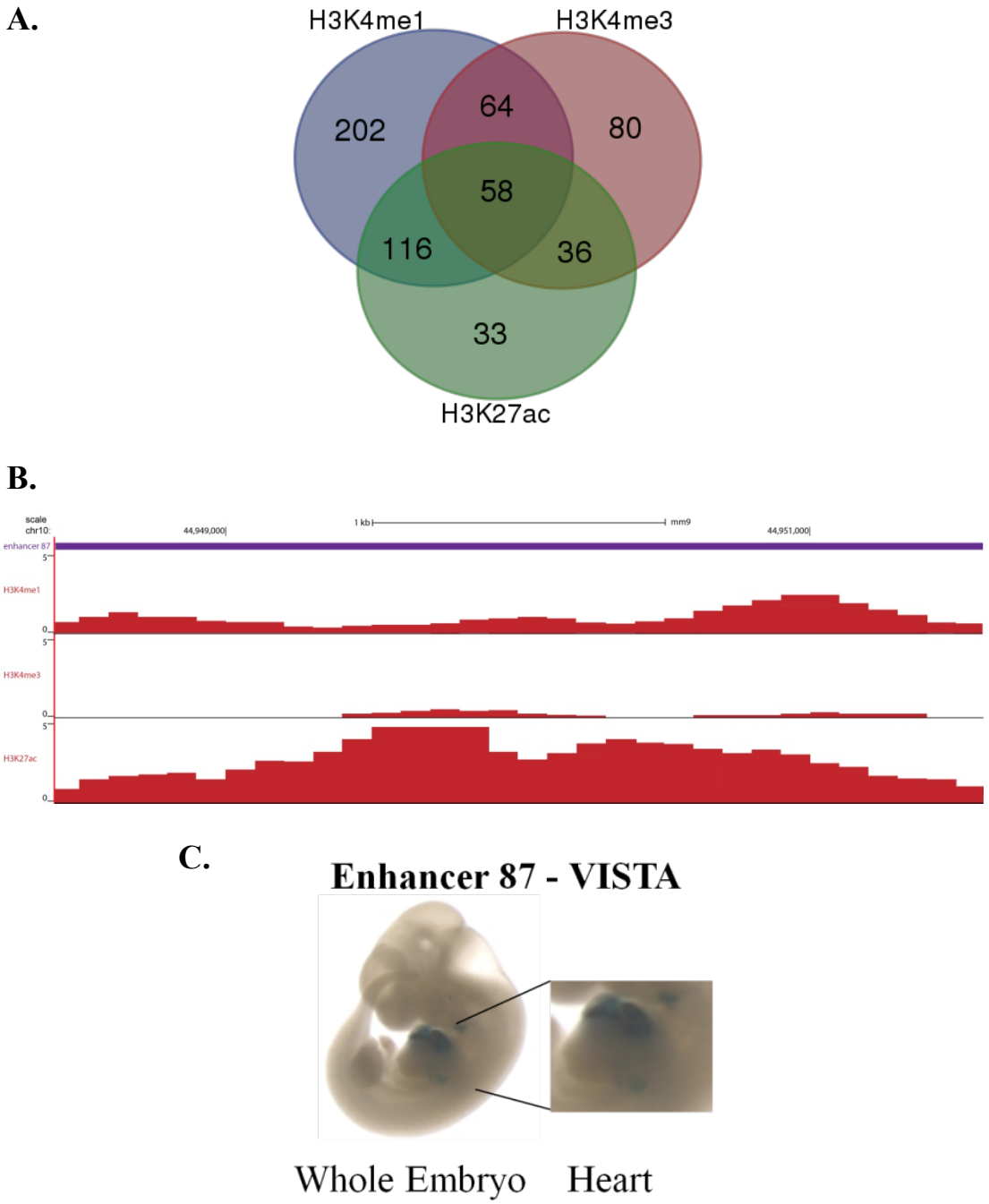


Figure 4.23: A fetal liver Rlf-DMR overlaps with an active enhancer in the fetal heart.

A. Venn diagram of E14.5 Rlf-DMRs that overlap with histone marks associated with regulatory elements in the heart (H3K4me1, enhancer mark; H3K27ac, active enhancer mark; H3K4me3, active promoter mark). **B.** UCSC Genome Browser tracks of a validated E11.5 enhancer mm87 (red bar) overlapping active enhancer marks H3K4me1 and H3K27ac, but not the promoter mark H3K4me3 on Chromosome 10. **C.** Transgenic mouse assay data from the VISTA database shows expression of enhancer mm87 is cardiac specific, (Visel et al., 2007).

I then asked whether any difference in methylation at this enhancer is observed in *Rlf* mutant fetal hearts at different stages in development. I undertook bisulphite sequencing of this enhancer in E13.5 and E18.5 whole hearts, as described previously. The *MommeD28* mutant line, in which GWBS was originally undertaken, was chosen for validation to minimise any differences that may occur between the different *Rlf* mutant strains. DNA from two *Rlf* $^{+/+}$ and two *Rlf* $^{MommeD28/MommeD28}$ E18.5 and E13.5 whole fetal hearts was converted and sequenced as described previously, refer **Section 3.2.4**.

At E13.5 a small increase in methylation, p <0.0001, was observed in *Rlf* $^{MommeD28/MommeD28}$ whole hearts (92% methylated) compared to *Rlf* $^{+/+}$ (77% methylated) whole hearts, **Figure 4.24A**. This difference in methylation between genotypes was maintained, p <0.0242, at E18.5 (*Rlf* $^{MommeD28/D28}$ – 93% methylated, *Rlf* $^{+/+}$ – 85% methylated), **Figure 4.24B**. At both time-points the percentage of methylation at this enhancer was relatively high (> 75%). As active enhancers have been commonly found to be hypomethylated this suggests this enhancer is not active at these time-points in this tissue (Hon et al., 2013a; Lister et al., 2009). Enhancer mm87 was identified by Visel et al., 2007, as being active at E11.5. Investigations in the future into methylation state at earlier time-points in the heart may reveal methylation of this enhancer to be time-point specific. Alternatively, the methylation observed in whole fetal hearts may not be illustrative of underrepresented cardiac cell types, such as endocardial cells in which *Rlf* mRNA expression has been observed. As such, undertaking bisulphite sequencing in whole fetal hearts may be diluting possible cell type specific methylation effects that are influencing the phenotype observed. This will be further discussed in **Section 4.3.**

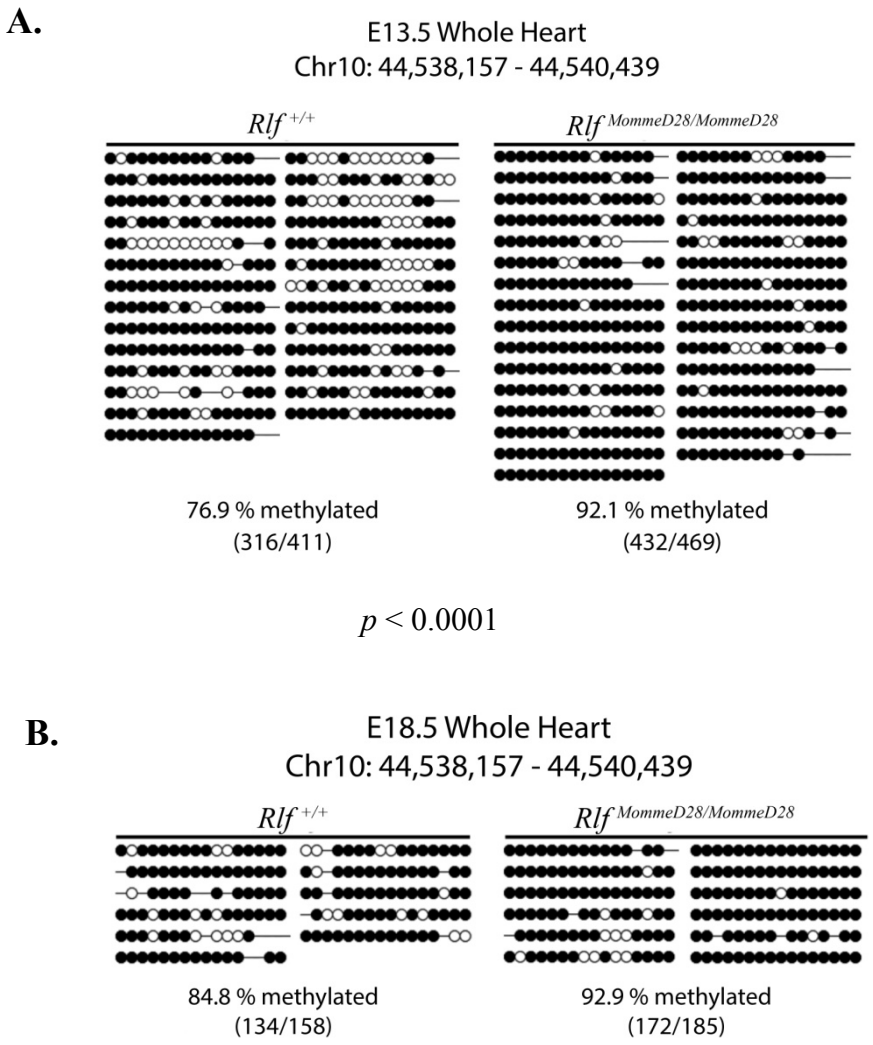


Figure 4.24: Methylation analysis of enhancer mm87 in *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ whole fetal heart.

A, B. Bisulphite sequencing of enhancer mm87 of whole E13.5 (A) and E18.5 (B) hearts reveals a significant difference in DNA methylation in *Rlf*$^{MommeD28/MommeD28}$ compared to wild-types (n = 2 per genotype). Each column represents DNA from a single heart, each row is the sequence from a single cell and each circle represents one CpG site. Filled circles represent methylated CpGs. Black circles represent methylated CpGs, white circles represent unmethylated CpGs, and lines (-) represents a non CpG or ambiguously sequenced position where a CpG exists in the genomic sequence. (X^2, $p < 0.0001$ A, $p < 0.05$ B).

4.3 DISCUSSION

Many groups have shown the importance of epigenetic modifiers in development. Studies involving a large number of different modifiers have shown that global deletion often results in embryonic or postnatal lethality (Ashe et al., 2008; Daxinger et al., 2013; Montgomery et al., 2007; Okano et al., 1999). The effect of loss of *Rlf* on viability and weight (**Table 4.1, Figure 4.8** and **Figure 4.9**) is reflective of results observed in other epigenetic mutant mouse lines both from the Whitelaw group and other research groups (Ashe et al., 2008; Daxinger et al., 2013; Lee et al., 2000). This result directly addresses the central hypothesis of this thesis; Rlf is required for normal development.

Here I have identified a putative heart defect in two different null *Rlf* mutant mouse lines and shown loss of Rlf results in perinatal lethality (**Table 4.1** and **Figure 4.10A-B**). Recent studies from our collaborator Dr G. Del Monte Nieto, VCCRI, have shown that *Rlf* mRNA is expressed in the endocardium and epicardium of the developing heart (Dr G. Del Monte Nieto, *personal communication*). The endocardium plays an important role in proper formation of the trabecular myocardium, whilst the epicardium contributes to multiple cardiac lineages, including smooth muscle cells and cardiac fibroblasts (Stankunas et al., 2008; Zhou et al., 2008). Non-compaction defects such as those observed in *Rlf* null mutants have been shown to arise from dysregulation within either the myocardial or endocardial layer (Chen et al., 2013; Luxan et al., 2013; Yang et al., 2012). Indeed, aberrant gene expression in the endocardium results in non-compaction phenotypes similar to those observed in *Rlf* mutant mice. For example, loss of *Fkbp1a* (FK506

binding protein 1a) in the endocardium resulted in ventricular hypertrabeculation and non-compaction, as a result of disrupted Notch1 signalling (Chen et al., 2013). Notch1 signalling has been found to be crucial for ventricular development, via mediating endocardial and myocardial interactions, and inactivation of Notch1 in the endocardium results in a non-compaction defect in mice (Grego-Bessa et al., 2007).

Following RNA-seq analysis of *Rlf* mutant hearts, no difference in *Notch1* expression was observed. This was not an unexpected result as Notch1 proteins are activated upon ligand binding (Grabher et al., 2006). Chen et al., 2013, found both over expression and loss of Fkbp1a had no impact on *Notch1* mRNA expression, but did alter Notch1 protein expression and mRNA expression of Notch1 target genes. Further investigation of gene expression in *Rlf* mutant hearts revealed a small but significant decrease in expression of Notch1 target genes (*Hey1, Erbb2*, and *Bmp2*, fold change 1.3, 1.2, and 1.2 respectively), as well as two Notch1 ligands (Jag1, fold change 1.2 and Jag2, fold change 1.3) (**Appendix X**). Analysis by our collaborator further found a reduction in activated Notch1 expression in *Rlf* mutant hearts, supporting the findings presented here (Dr G Del Monte Nieto, *personal communication*).

A recent publication found the Notch ligands *Jag1* and *Jag2* play important roles in trabecular compaction and the maturation of cardiac chamber development (D'Amato et al., 2016). The authors found loss of *Jag1* or *Jag2* in the mouse resulted in a phenotype reminiscent of human LVNC, similar to that observed in *Rlf* mutants. It could be proposed that loss of Rlf results in down regulation of the Jagged ligands

which in turn results in lost Notch1 activation leading to the phenotype observed in *Rlf* mutants. However, further investigation is needed to determine the specific mechanisms by which loss of *Rlf* impacts the Notch pathway.

While a putative heart defect has been observed in these mice it is important to note that the placenta also plays an important role in regulating heart development (Barak et al., 1999). While the presence of heart defect may be the result of placental defects, this is not always the case (Wang et al., 2004). Barak et al., 1999, investigated *Ppar-γ,* peroxisome proliferator–activated receptor γ, and found loss of this gene resulted in defects in placental vascularisation, thinning of the myocardium in the heart and death at E10. Aggregation of mutant *Ppar-γ* embryos with wild-type placentas found the cardiac defect was corrected when no placental defect was present. In knockout mice of the chromatin remodeler *Pbrm1,* Polybromo 1, Wang et al., 2004, observed defects compromising fetal-maternal exchange in the placenta, incomplete development of the ventricle walls and death between E12 – E15. However, unlike Barak et al.'s, 1999, study, when knockout *Pbrm1* embryos were aggregated with wild-type placentas normal heart development was not restored and the cardiac defect remained in these embryos. These embryos were sacrificed at mid-gestation (E14.5) and the impact of the cardiac defect on development or neonatal survival not investigated (Wang et al., 2004).

In this study I analysed placentas from wild-type and homozygous *MommeD28* E14.5 embryos and found subtle differences in fetal placenta vasculature between the two genotypes (**Figure 4.15** and **Figure 4.16**). Due to time and financial constraints,

no further functional, transcriptomic or histological studies were undertaken in the *Rlf* mutant mice. Defects in placenta vascular morphology, such as that observed in *Rlf* $^{MommeD28/MommeD28}$ placentas can decrease the surface area available for nutrient, oxygen and waste exchange (Gong et al., 2011). Additionally, fetal growth has been shown to be closely linked to this exchange, and deficiency in exchange can result in intrauterine growth restriction, cardiovascular defects and perinatal morbidity (Fowden et al., 2008). As *Rlf* $^{MommeD28/MommeD28}$ mice have been found to be significantly smaller at mid- and late gestation, the defect in placenta vasculature observed at E14.5 may impact on fetal growth in these mice. Undertaking embryo aggregation (aggregating homozygous *Rlf* embryos with wild-type placentas), as performed in the studies by Barak et al., 1999, and Wang et al., 2004, would be the most beneficial in determining whether the placental vascular defect present in *Rlf* mutants is the cause of or independent to the defects in cardiac development and fetal growth observed.

The cardiac and placental morphology presented here give rise to alternate theories as to how the phenotype may be altered in *Rlf* null mice. The cardiac defect identified in *Rlf* null mice may be the result of aberrant gene expression in the endocardium, as a consequence of loss of *Rlf,* leading to the non-compaction phenotype observed. The differences in placental vasculature may be preventing appropriate cardiac and fetal development as a consequence of insufficient maternal and fetal exchange. Alternatively, the phenotype observed in *Rlf* null mice could be due to differences in both the placenta and cardiac development. The development of conditional knockout mouse models, or producing chimeras by aggregating *Rlf* null embryos with wild-type placentas, will help to address these questions in the future.

Previously, tsDMRs have been shown to mark enhancers that were once active at early stages of development. These tsDMRs retain their hypomethylated state in adult tissue but active histone marks once present at earlier time-points are no longer present (Hon et al., 2013a). From the Rlf-DMRs identified in the fetal liver some were also found to overlap with sites enriched for active histone marks in the fetal heart (**Figure 4.23**). Here I investigated one of these Rlf-DMRs in E13.5 and E18.5 whole hearts as it had already been validated as an active enhancer in E11.5 hearts and overlapped with active histone marks in E14.5 fetal heart (**Figure 4.23**) (Visel et al., 2007). A significant increase in methylation was observed comparing the two genotypes, regardless of the already highly methylated state in wild-type hearts at both E13.5 and E18.5 (77% and 85% respectively) **(Figure 4.24B-C**). Interestingly, this statistical significance decreased with the age of pups (E13.5 $p<0.0001$; E18.5 $p<0.05$). Further investigation of this enhancer at E11.5, the time point in which it was identified at Visel et al., 2007, may reveal a greater change in methylation between genotypes, pointing towards this enhancer being active early in the developing heart.

Alternatively, undertaking methylation analysis in a single cardiac cell type population may be a superior method and yield greater information. *Rlf* has been found to be expressed in the endocardium and epicardium of the developing heart (Dr G. Del Monte Nieto, *personal communication*). These two cell types make up a small percentage of the total cell population within the heart (~15%) (Banerjee et al., 2007). Investigation of a single cardiac cell type population may reveal these methylation changes are cell type specific and may elucidate potential mechanism through which loss of *Rlf* results in the phenotype observed.

4.4 FUTURE DIRECTIONS

This is the first study investigating *Rlf* and the effect of loss of *Rlf* on phenotype. Here we identified a putative heart defect in mice lacking *Rlf* expression.

In Australia congenital heart defects account for 20% of perinatal deaths, still births or death within the first 28 days of life (Australian Institute of Health and Welfare, 2011). However the genetic cause for many of these cases is unknown. As the heart defect identified in *Rlf* mutant mice is similar to left ventricular non-compaction (LVNC) observed in humans, undertaking genetic sequencing for *Rlf* will determine whether mutations in *Rlf* are also present in humans with the same defect. A collaboration has been established with Professors Chris Semsarian (University of Sydney) and Jose de la Pompa, (Centro Nacional de Investigaciones Cardiovasculares, Spain), who both oversee large collections of clinical LVNC samples. Undertaking sequencing of these clinical samples for *RLF* may enhance the current genetic screening, diagnosis and management strategies for patients with LVNC.

Undertaking detailed analysis of cardiac function and structure of either *Rlf Momme* mutants or a conditional cardiac *Rlf* knockout in early neonatal or late postnatal pups would be beneficial. Results from this study suggest that homozygous *Rlf* null mice ($Rlf^{MommeD28/MommeD28}$ and $Rlf^{MommeD34/MommeD34}$) are likely to die shortly after birth, however no investigation into the impact of this defect on cardiac function was undertaken. The studies by Montgomery et al., 2007, and Wang et al.,

2004, which investigated the impacts of loss of the epigenetic modifiers *Hdac2* and *Pbrm1,* found loss of either modifier resulted in decreases in heart rate prior to death. Therefore, determining whether *Rlf* mutants have altered cardiac function prior to or after birth may help to define the role of *Rlf* in the heart. Furthermore, while the majority of homozygous *MommeD28* and *MommeD34* mutants die shortly after birth, a small number survive to weaning. Identifying clearly how loss of *Rlf* alters cardiac structure and function in the mouse may identify strategies with which to overcome this lethality.

Further analysis of the placental defects observed will be undertaken (dependent on funding) in the future in collaboration with Dr David Simmons (University of Queensland). Investigating markers specific to different placenta cell types, such as *Gcm1* (glial cells missing homolog 1) and *Syna* (Syncytin a), markers specific to fetal and maternal blood vessels respectively, may help to define the role of *Rlf* in vascular development in the placenta. Due to the collapsed appearance of fetal vessels in *Rlf* mutant placentas utilising stereology to define the structural and spatial arrangements within the whole placenta, and the individual placenta layers, may more accurately define the phenotype observed in these placentas.

Co-immunoprecipitation experiments (co-IP) experiments performed by the Whitelaw laboratory have identified putative binding partners of Rlf *in-vitro* that have established roles in transcription and chromatin modification, or in replication and DNA repair (Harten et al., 2015). Undertaking co-IP of wild-type and homozygous *Rlf* hearts followed by Mass spectrometry identification of protein binding partners would help to define the role Rlf plays in the developing heart. In

order to be able to gain significant results, hearts of late gestational embryos would be the best developmental time-point for co-IP investigation. A greater amount of protein can be extracted from the larger, late gestation hearts; reducing the number of embryos from across different litters that would otherwise need to be collected at earlier developmental stage and reducing inter- litter variation. The identification of Rlf binding partners' *in-vivo* may identify protein binding partners that are critical in cardiac development.

Chapter 5: Investigating the consequence of reduced *Rlf* expression in adult mice

5.1 INTRODUCTION

In contrast to complete absence of an epigenetic modifier, many studies have shown mice with reduced levels of a particular epigenetic modifier are viable, fertile and phenotypically normal (Okano et al., 1999; Stopka & Skoultchi, 2003). For this reason, few studies focus on the impact of heterozygosity or reduced dosage of epigenetic modifiers on phenotype. The *MommeD28* and *MommeD34 Rlf* null lines do not produce viable homozygous mutants at three weeks of age, however the hypomorphic *MommeD8* line does produce homozygous offspring that are viable and fertile (Daxinger et al., 2013). In this chapter two hypomorphic *Rlf* mouse mutants, $Rlf^{MommeD8/MommeD34}$ and $Rlf^{MommeD8/MommeD8}$, which have reduced *Rlf* expression, will be used to determine whether differences in *Rlf* expression impacts upon phenotype, viability and fertility.

5.1.1 Reduced dosage of epigenetic modifiers impacts on viability

Mice with reduced levels of epigenetic modifiers display similarities in their effects on viability and phenotype to that observed in *Rlf* mutant mouse lines. Heterozygous mutants from a number of mouse lines identified in the *Momme* screen are viable and display no overt changes in phenotype. For example, *Hdac1* and *Smarca5* heterozygous mutant mice, but not homozygous mutants, are viable at weaning but weigh less than their wild-type littermates (Ashe et al., 2008; Chong et

al., 2007). This is similar to observations made in the *MommeD28* and *MommeD34* null *Rlf* lines (this study and Daxinger et al., 2013). Mice with hypomorphic mutations have been shown to have varying impacts on phenotype. Mice carrying two hypomorphic *Dnmt3b* alleles are viable at weaning, although not at the expected Mendelian ratio, and are fertile. Mice with two hypomorphic *Setdb1* (SET domain, bifurcated 1) alleles also survive to adulthood but have decreased fertility (Daxinger et al., 2013; Youngson et al., 2013). These findings are similar to the hypomorphic *MommeD8* mouse line. Whilst the viability of homozygous *MommeD8* mice has previously been addressed, no further research investigating *Rlf* $^{MommeD8/MommeD8}$ or Rlf $^{MommeD8/MommeD34}$ mice has been undertaken.

5.1.2 Epigenetics and ageing

Ageing is a result of decline in regeneration and reproduction, and as a result of this decline an organism may become more susceptible to disease, stress and injury (Moskalev et al., 2014). Recently, studies in the field of ageing have identified gerontogenes, genes that control ageing and longevity (Moskalev et al., 2014). Additionally, epigenetic modifiers have been found to play key roles in the ageing process. For example, mice deficient for the histone deacetylase *Sirt6*, sirtuin 6, exhibit premature signs of ageing, including sudden loss of subcutaneous fat and bone density, curvature of the spine, and premature death at approximately three weeks of age (Mostoslavsky et al., 2006). Whilst mice over- expressing *Sirt6* were found to have an increased lifespan of approximately 15% in males compared to wild-type littermates, no difference was observed in females over-expressing *Sirt6* (Kanfi et al., 2012). No previous studies have investigated how reduced *Rlf* expression may impact upon ageing or onset of adult disease in the mouse.

5.1.3 Transgenerational epigenetics

Inbred mouse colonies make it possible to study transgenerational epigenetic inheritance in mice that are genetically identical. The viability of *Rlf*$^{MommeD8/MommeD8}$ mice, and their ability to reproduce successfully, presents a unique opportunity to examine the effects of reduced *Rlf* expression across multiple generations. Such studies have been performed in hypomorphic *Dnmt3b* mice. Investigating the effect of reduced *Dnmt3b* expression across generations found a recovery in bodyweight in F2 offspring versus F1 offspring. No other differences in weight were observed across other generations (Youngson et al., 2013). As *Dnmt3b* is required for *de novo* DNA methylation, Youngson et al., 2013, also investigated the impact of reduced *Dnmt3b* expression on methylation of the X-linked gene *Hprt.* No cumulative effect was observed in late generation *Dnmt3b* offspring, however the authors do not rule out a role for *Dnmt3b* in transgenerational epigenetic inheritance.

Both *Rlf* and *Dnmt3b* expression has been observed in the zygote, two cell and four cell embryo using single cell RNA-seq profiling (Yan et al., 2013). These are cell types in which epigenetic reprograming is occurring following fertilisation of the zygote, with the maternal and paternal genome are undergoing passive and active DNA demethylation, respectively (Smallwood & Kelsey, 2012). The presence of *Rlf* expression at this developmental time-point suggests it may play a role in epigenetic reprogramming in the mouse.

As loss of *Rlf* has previously been shown to alter bodyweight and methylation in the different *Rlf Momme* lines, investigating the role of *Rlf* across generations may reveal a role for *Rlf* in inheritance of epigenetic state across generations.

5.2 RESULTS

5.2.1 Presence of the *MommeD8* allele alters gene expression in the fetal liver

Previous experiments by the Whitelaw lab have shown Rlf protein expression is reduced in $Rlf^{MommeD8/MommeD8}$ fetal head lysates (Daxinger et al., 2013). Loss of *Rlf* has also been shown to alter gene expression in both *MommeD28* and *MommeD34* homozygous E14.5 liver (**Section 3.2.2**). Heterozygosity for the *Rlf MommeD34* allele was also shown to result in significant changes in gene expression. Following on from these results, I was interested in determining whether reduced dosage of *Rlf,* as observed in homozygous *MommeD8* mice, could have a similar effect. Analysis of four up- and four down-regulated genes analysed in **Section 3.2.2** was undertaken using $Rlf^{+/+}$ and $Rlf^{MommeD8/MommeD8}$ E14.5 liver (n = 5 per genotype). A significant difference in gene expression, in all but one gene, was observed between the two genotypes following qRT-PCR, **Figure 5.1.** This data shows that reduced expression of *Rlf* in $Rlf^{MommeD8/MommeD8}$ mice is capable of impacting gene expression. Whilst no putative Rlf binding sites, following ChIPseq in the fetal liver, were found to overlap differentially expressed genes, whether their change in expression is directly due to loss of *Rlf* is still unknown.

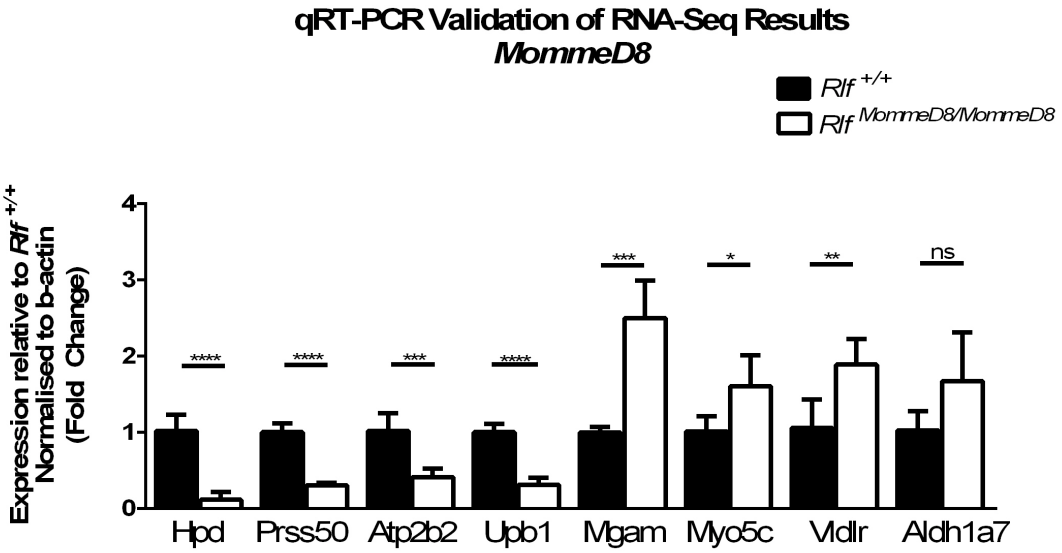
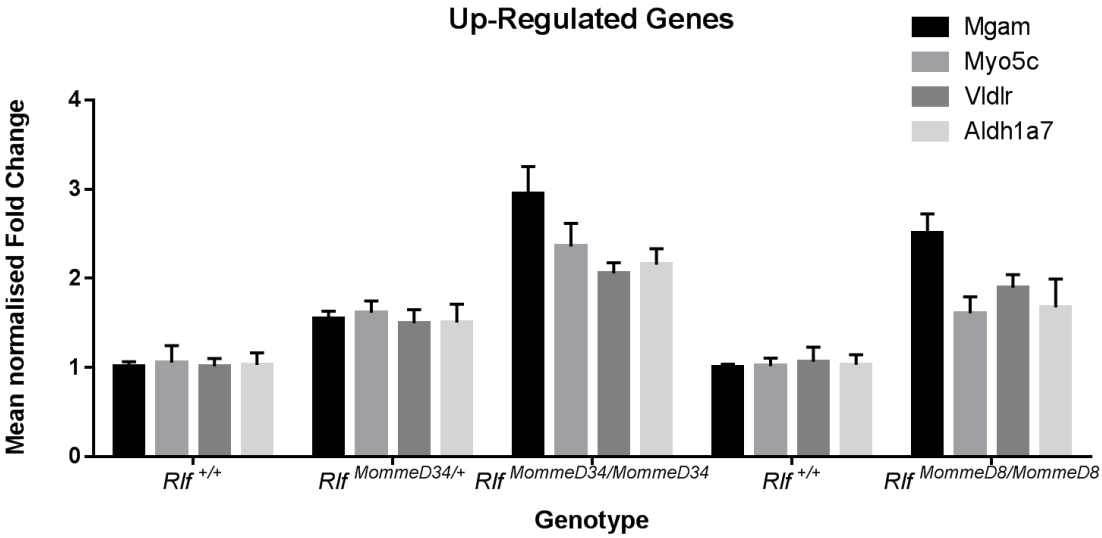


Figure 5.1: The hypomorphic *MommeD8* allele alters expression of some genes in the fetal liver.

qRT-PCR analysis of eight significantly differentially expressed genes identified from RNA-seq show all but one to also be differentially expressed in *Rlf* $^{MommeD8/MommeD8}$ E14.5 liver. mRNA expression was normalised to β-actin and wild-type controls. Error bars represent SEM and Student's t-test was used to calculate *p* values (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, ****$p < 0.0001$). $n = 5$ for each genotype.

As *Rlf* $^{MommwD8/MommeD8}$ mice still produce some Rlf protein (hypomorphic allele), whereas *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ mice do not produce any Rlf protein (null alleles), I was interested in determining whether *Rlf* $^{MommwD8/MommeD8}$ mice have similar gene expression patterns to *Rlf* $^{MommeD34/+}$ mice. I chose the *MommeD34* line as it had previously been used for qRT-PCR validation of RNA-seq results. The mean normalized fold change in expression following qRT-PCR was used for comparison between the *MommeD8* and *MommeD34* lines. The mean normalized fold change of gene expression in *Rlf* $^{MommeD8/MommeD8}$ mice was observed to be slightly lower (for up-regulated genes) than *Rlf* $^{MommeD34/MommeD34}$ mice, but higher than *Rlf* $^{MommeD34/+}$ mice, **Figure 5.2A**. The opposite was observed when analysing the down- regulated genes, **Figure 5.2B**. This suggests that carrying two of the hypomorphic *Rlf MommeD8* alleles has a greater impact on endogenous gene expression than carrying one copy of the *Rlf MommeD34* allele and one wild-type *Rlf* allele.

A.



B.

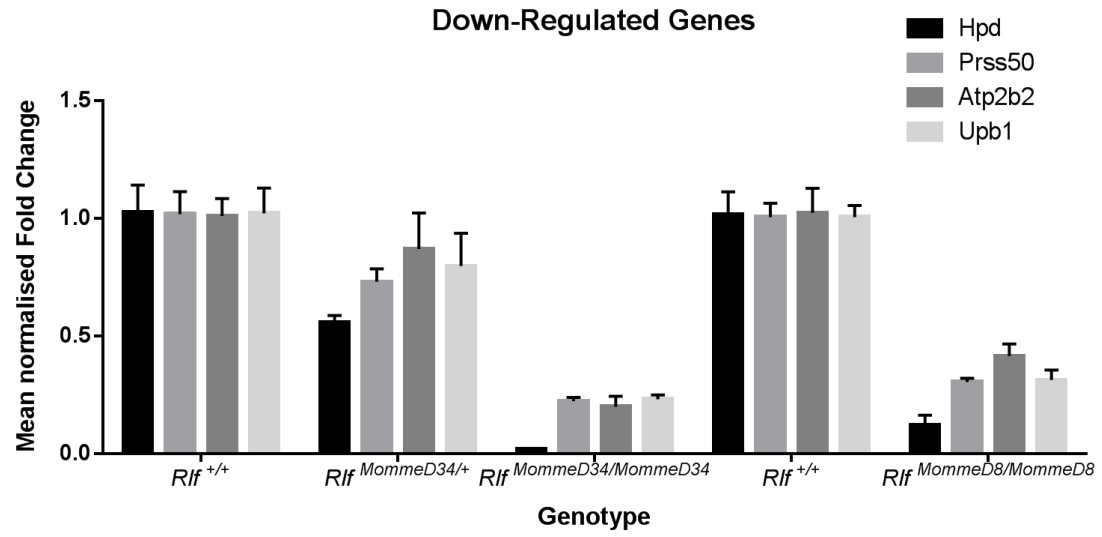


Figure 5.2: Endogenous gene expression is reduced in *Rlf*$^{MommeD8/MommeD8}$ fetal liver more than *Rlf*$^{MommeD34/+}$ fetal liver.

A, B. Mean normalised fold change comparison of up- and down-regulated genes in the *MommeD34* and *MommeD8* mouse lines reveals a stronger change in *Rlf* $^{MommeD8/MommeD8}$ compared to *Rlf* $^{MommeD34/+}$ mice but not as strong as *Rlf* $^{MommeD34/MommeD34}$ mice. Error bars represent SEM. *n* = minimum of 4 per genotype.

5.2.2 Western blotting and gene expression analysis of tissue extracted from compound heterozygous (*Rlf*$^{MommeD8/MommeD34}$) mice

Previously it has been shown that Rlf protein expression is reduced in *Rlf* $^{MommeD8/MommeD8}$ mice and not present in *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ mice (Daxinger et al., 2013). However, no experiments have determined the level of Rlf expression in mice compound heterozygous for different *Rlf Momme* alleles, or the impact of compound heterozygosity on endogenous gene expression.

Compound heterozygous *Rlf* $^{MommeD8/MommeD34}$ mice are predicted to have low protein activity, as a result of the presence of a null allele (*MommeD34*) and a hypomorphic allele (*MommeD8*). To assess this *Rlf*$^{MommeD8/MommeD34}$ mice (bred from a *Rlf* $^{MommeD8/MommeD8}$ x *Rlf* $^{MommeD34/+}$ cross) and age matched wild-type adult mice were used for Western blotting analysis, and no overt change in Rlf protein level was observed (data not shown). Reasons for the lack of change in Rlf protein expression in these mice are currently unclear.

Next, I asked whether expression of *Hpd* mRNA, a gene previously found to be significantly differentially expressed in fetal liver from both of the *Rlf* null *Momme* lines (refer **Figure 3.**), and also hypomorphic *Rlf* $^{MommeD8/MommeD8}$ mice (refer **Figure 5.1**). mRNA was collected from fetal liver from *Rlf* $^{MommeD8/+}$ and *Rlf* $^{MommeD8/MommeD34}$ littermates (*n* = minimum 3 per genotype), and also *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ littermates (*n* = 3 per genotype). A significant decrease in *Hpd* mRNA expression was observed in *Rlf* $^{MommeD28/MommeD28}$, but not *Rlf* $^{MommeD28/+}$ E14.5 liver, compared to

wild-type littermates, Figure 5.3. The expression of *Hpd* mRNA in *Rlf* $^{MommeD8/+}$ mice was similar to that observed in *Rlf* $^{+/+}$ mice, whilst a significant decrease in expression was observed in *Rlf* $^{MommeD8/MommeD34}$ mice compared to *Rlf* $^{MommeD8/+}$ littermates, Figure 5.3.

The results obtained here in *Rlf* $^{MommeD8/MommeD34}$ offspring reflect those observed in the *Rlf* $^{MommeD8/MommeD8}$ mice, with the presence of two hypomorphic alleles having a greater impact on gene expression than presence of one *Rlf* null and one *Rlf* wild-type allele.

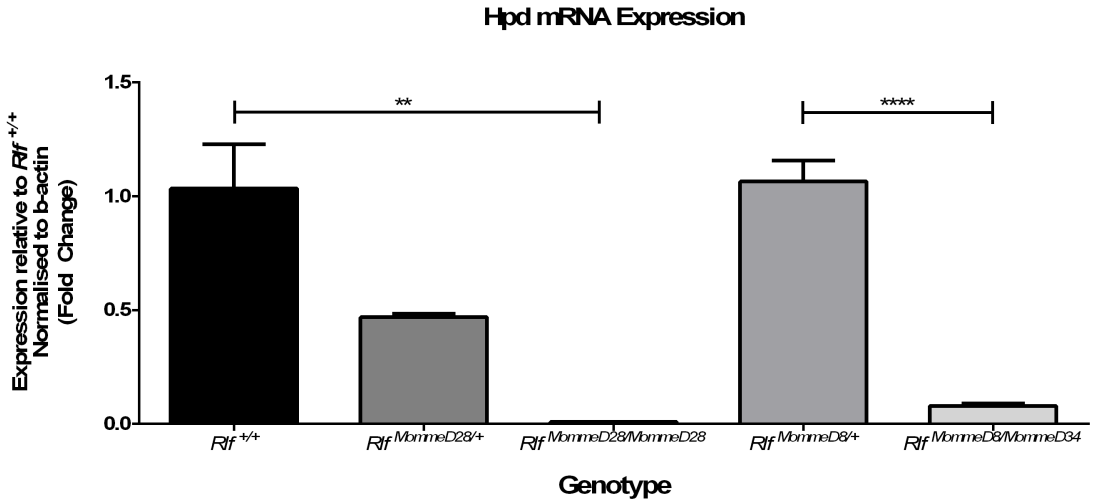


Figure 5.3: qRT-PCR of *Rlf* $^{MommeD8/MommeD34}$ mice.

qRT-PCR analysis of *Hpd* identified as being differentially expressed from RNA-seq data was performed on RNA extracted from the *Rlf* $^{+/+}$, *Rlf* $^{MommeD28/+}$, *Rlf* $^{MommeD28/MommeD28}$, *Rlf* $^{MommeD8/+}$, and *Rlf* $^{MommeD8/MommeD34}$ E14.5 liver. mRNA expression was normalised to β-actin and wild-type controls. Error bars represent SEM and Student's t-test was used to calculate *p* values (**$p < 0.005$, ****$p < 0.0001$). n = min. 3 per genotype.

5.2.3 Predicted *Rlf* activity in the different mutant mouse lines

As differences in Rlf protein expression have been observed amongst the different *Rlf Momme* mutant lines, I wanted to try to define how the activity of Rlf compares across each of the different mutant mouse lines. To do this I investigated endogenous *Hpd* mRNA expression in each of the *Rlf Momme* lines collected during this PhD, and compared these to previously published data investigating Rlf protein expression from Western blotting, and GFP expression in the independent *Rlf* lines.

Hpd mRNA expression, quantified from qRT-PCR performed in this PhD showed a reduction in *Hpd* expression in each of the mutant lines compared to wild-type mice, **Table 5.1**. Expression was reduced the most in *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ mice (5%), followed by *Rlf* $^{MommeD8/D34}$ (10%), *Rlf* $^{MommeD8/D8}$ (20%), *Rlf* $^{MommeD28/+}$ (40%), *Rlf* $^{MommeD34/+}$ (60%) and *Rlf* $^{MommeD8/+}$ mice (100% expression, **Table 5.1**. Furthermore, in lines where multiple genes were analysed the effect on expression was consistent across the genes analysed. The differences in *Hpd* mRNA expression observed correlates with previous findings investigating GFP and Rlf expression in the *Momme* lines. Combining all of these characteristics it can be predicted that the activity of Rlf is most severely altered in *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ mice (no activity), followed by *Rlf* $^{MommeD8/MommeD34}$ and *Rlf* $^{MommeD8/MommeD8}$ (low activity), *Rlf* $^{MommeD28/+}$, *Rlf* $^{MommeD34/+}$ (moderate activity), and *Rlf* $^{MommeD8/+}$ mice (complete activity) (**Table 5.1**).

Table 5.1: Predicted activity of Rlf in different mutant mouse line

Genotype	*Hpd* mRNA expression	GFP expression	Protein Expression	Predicted Rlf activity
Rlf $^{+/+}$	100% [2,5,6]	55% [1]	100% [1]	Complete activity
Rlf $^{MommeD8/+}$	100% [5]	41% [3]	100% [1]	Complete activity
Rlf $^{MommeD28/+}$	40% [2]	44% [1]	50% [1]	Moderate activity
Rlf $^{MommeD34/+}$	60% [6]	39% [1]	50% [1]	Moderate activity
Rlf $^{MommeD8/MommeD8}$	20% [5]	15% [3]	50% [1]	Low activity
Rlf $^{MommeD8/MommeD34}$	10% [2]	13% [4]	100% [2]	Low activity
Rlf $^{MommeD34/MommeD34}$	5% [6]	11% [1]	0% [1]	No activity
Rlf $^{MommeD28/MommeD28}$	5% [6]	7% [1]	0% [1]	No activity

Table 5.1: Predicted Rlf protein activity in each of the *Momme* mouse lines, estimated from published data and data from this PhD. Rlf Protein and GFP expression were estimated from previously published Western blots and Western blots performed during this PhD, *Hpd* mRNA expression was calculated for each of the *Rlf Momme* lines in Chapters 3 and 5 of this PhD, and together this information was used to predict Rlf activity. [1] Daxinger et al., 2013; [2] Figure 5.3, this study; [3] Ashe et al., 2008; [4] Figure 5.4, this study; [5] Figure 5.1, this study; [6] Figure 3.2, this study.

5.2.4 Reduced dosage of *Rlf* alters viability and weight of mice

Whilst viability in each of the *Rlf* null lines (no Rlf activity) has been addressed the effect of reduced dosage on viability has not. *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ mice have been found to die shortly after birth while *Rlf* $^{MommeD8/MommeD8}$ offspring survive in less than expected numbers (**Table 4.1** and Daxinger et al., 2013). No previous work has investigated the viability of *Rlf* $^{MommeD8/MommeD34}$ mice.

To determine whether *Rlf* $^{MommeD8/MommeD34}$ mice were viable after birth, I crossed a *MommeD8* homozygous mouse to a *MommeD34* heterozygous mouse to produce *Rlf* $^{MommeD8/+}$ and *Rlf* $^{MommeD8/MommeD34}$ (compound heterozygous) offspring. Homozygous *MommeD8* mice were used to increase the likelihood of breeding mice carrying both the *MommeD8* and *MommeD34* allele. Almost half of the expected number of *Rlf* $^{MommeD8/MommeD34}$ offspring was observed at weaning (expected 50%, observed 30%), **Figure 5.4A**. I then wanted to determine whether hypomorphic *Rlf* $^{MommeD8/MommeD34}$ mice had an altered GFP expression similar to the other *Rlf Momme* mouse lines. In *Rlf* $^{MommeD8/MommeD34}$ mice the percentage of GFP expressing erythrocytes was markedly reduced (13% ± 2) compared to their *Rlf* $^{MommeD8/+}$ littermates (41% ± 5), **Figure 5.4B**. The difference in GFP expression in hypomorphic *Rlf* $^{MommeD8/MommeD34}$ mice reflects that observed in *Rlf* $^{MommeD8/MommeD8}$ hypomorphic mice (Daxinger et al., 2013). These results suggests carrying both the *MommeD8* and *MommeD34* alleles has a greater impact on viability than carrying one copy of the *MommeD8* allele and a wild-type allele, and correlates with the predicted activity of Rlf in these mice.

As reduced numbers of *Rlf* $^{MommeD8/MommeD34}$ offspring were observed at three weeks of age, I asked whether this impacted upon litter size. No statistical difference in litter size between pups born from a wild-type cross or *Rlf* $^{MommeD8/D8}$ x *Rlf* $^{MommeD34/+}$ cross was observed, **Figure 5.4C**. Although a trend towards decreased litter size was observed with litters born from a *Rlf* $^{MommeD8/MommeD8}$ x *Rlf* $^{MommeD34/+}$ containing less mice (average 5.9 pups) compared to those born from wild-type parents (average 7.3 pups). This is consistent with the decrease in viability of *Rlf* $^{MommeD8/MommeD34}$ offspring observed.

Hypomorphic *Rlf* $^{MommeD8/MommeD8}$ offspring have been previously found to weigh less than their wild-type litter mates at weaning (Ashe et al., 2008). As a result of this I wanted to determine whether hypomorphic *Rlf* $^{MommeD8/MommeD34}$ mice were also reduced in weight at weaning. Although a large variation in pup weight was observed in both genotypes, *Rlf* $^{MommeD8/MommeD34}$ offspring were found to weigh significantly less (8.85 g, ± 0.604) than their *Rlf* $^{MommeD8/+}$ littermates (11.06 g, ± 0.30) at three weeks of age, **Figure 5.4D**. The reduced weight and viability observed in compound heterozygous offspring suggests reduced dosage of Rlf plays an important role in development and survival.

In the following sections hypomorphic *Rlf* $^{MommeD8/MommeD34}$ mice will be used in a series of preliminary experiments investigating the effect of reduced *Rlf* expression on longevity, ovarian morphology and cardiac structure and function.

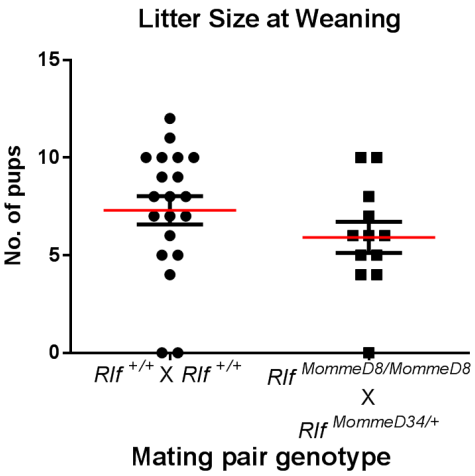
A.

Viability	Age	*Rlf* $^{MommeD8/+}$	*Rlf* $^{MommeD8/MommeD34}$
Rlf $^{MommeD8/MommeD8}$ X *Rlf* $^{MommeD34/+}$	3 weeks	51 (72%)	20 (28%)

B.

GFP Expressing cells	*Rlf* $^{MommeD8/+}$	*Rlf* $^{MommeD8/MommeD34}$
Rlf $^{MommeD8/MommeD8}$ X *Rlf* $^{MommeD34/+}$	41% ±5	13% ±2

C.



D.

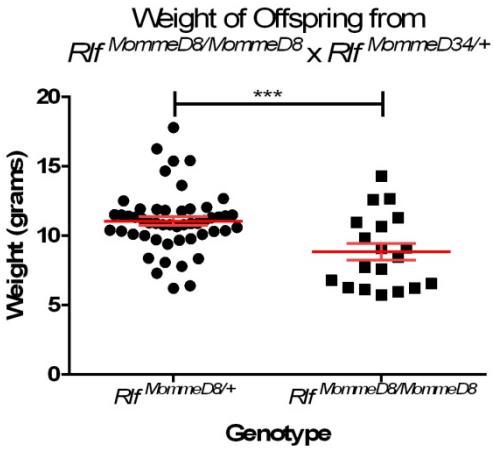


Figure 5.4: *Rlf* $^{MommeD8/MommeD34}$ offspring are smaller and less viable than *Rlf* $^{MommeD8/+}$ offspring.

A. Fewer *Rlf* $^{MommeD8/MommeD34}$ offspring from a *Rlf* $^{MommeD8/MommeD8}$ x *Rlf* $^{MommeD34/+}$ mating are observed at three weeks (weaning) compared to heterozygous *Rlf* $^{MommeD8/+}$ offspring. Tabulated data shows the number of observed mice and (% of total). B. Percentage of GFP expressing erythrocytes in *Rlf* $^{MommeD8/+}$ and *Rlf* $^{MommeD8/MommeD8}$ littermates at three weeks of age (mean ± SEM). C. No difference is observed when comparing litter sizes from a *Rlf* $^{MommeD8/MommeD8}$ x *Rlf* $^{MommeD34/+}$ cross to litter sizes born from a *Rlf* $^{+/+}$ intercross (n = 14 and 21 litters respectively). Each data point represents an independent litter, where the number of pups equals zero this refers to litters observed at birth but no longer present at weaning. D. Weights of offspring at three weeks showing *Rlf* $^{MommeD8/MommeD34}$ offspring are significantly smaller than their *Rlf* $^{MommeD8/+}$ littermates. Each data point represents an individual embryo. Error bars represent SEM and Student's t-test was used to calculate p values (***$p < 0.001$), n = 8 litters.

5.2.5 Reduced dosage of Rlf has no major impact on longevity

The compound heterozygous mice bred in **Section 5.2.4** were then used to study the effects of reduced dosage of Rlf on longevity. The cohort consisted of six male and four female $Rlf^{MommeD8/MommeD34}$ mice (generated from a $Rlf^{MommeD8/MommeD8}$ x $Rlf^{MommeD34/+}$ cross), and eight male and four female age matched wild-type mice (generated from a $Rlf^{+/+}$ x $Rlf^{+/+}$ cross). **Table 5.2** shows the dates of birth, weaning and death, and sex for each of these mice. All mice were housed at 21 – 23 °C with a 12 hour light and dark cycle and unlimited access to food and water.

From the male mouse cohort one $Rlf^{MommeD8/MommeD34}$ was found dead in cage at 21 months of age, an autopsy found this male had an inflated bladder but no other abnormalities (data not shown). A second $Rlf^{MommeD8/MommeD34}$ male was culled, at 17 months of age, due to signs of physical distress, presenting with a distended stomach. An autopsy of this male revealed this mouse also had an inflated bladder (data not shown). The findings in these males are also commonly observed in aged wild-type FVB/NJ mice maintained by our laboratory (Dr S. Harten, *personal communication*). As none of the other males in the cohort displayed any overt signs of distress or atypical behaviour, no further investigation into this phenotype was undertaken. The remaining mice were then used to study the effect of reduced dosage of *Rlf* on heart structure and function in **Section 5.2.6.**

Next Page:
Table 5.2: Compound heterozygous and wild-type mice bred for longevity study. $Rlf^{MommeD8/MommeD34}$ males (pale blue), $Rlf^{+/+}$ males (dark blue), $Rlf^{MommeD8/MommeD34}$ females (pink), $Rlf^{+/+}$ females (red). Date of birth (DOB), date of weaning (DOW), dead in cage (DIC), age at death in months (mths), and ECG (echocardiogram).

Table 5.2: Longevity study cohort of *Rlf*$^{MommeD8/MommeD34}$ and *Rlf*$^{+/+}$.

Genotype	ID #	Sex	DOB	DOW	Approx. Age at Death	Notes
Rlf$^{MommeD8/MommeD34}$	14.5.10	M	19.4.13	10.5.13	21mths	Found DIC 12/01/2015, inflated bladder observed
Rlf$^{MommeD8/MommeD34}$	14.5.20	M	13.7.13	02.8.13	20 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{MommeD8/MommeD34}$	23.12.27	M	16.9.13	11.10.13	17 mths	Culled 30/01/15, due to distended stomach caused by inflated bladder
Rlf$^{MommeD8/MommeD34}$	23.12.30	M	16.9.13	11.10.13	18 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{MommeD8/MommeD34}$	23.12.32	M	16.9.13	11.10.13	18 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	52	M	03.7.13	22.7.13	20 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	34	M	10.8.13	30.8.13	20 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	35	M	10.8.13	30.8.13	18 mths	Culled 30/01/15, for histology match to *Rlf*$^{MommeD8/MommeD34}$ 23.12.27
Rlf$^{+/+}$	36	M	10.8.13	30.8.13	19 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	37	M	10.8.13	30.8.13	18 mths	Culled 30/01/15, for histology match *Rlf*$^{MommeD8/MommeD34}$ 23.12.27
Rlf$^{+/+}$	20	M	15.8.13	11.9.13	19 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	21	M	15.8.13	11.9.13	19 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	22	M	15.8.13	11.9.13	19 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{+/+}$	23	M	15.8.13	11.9.13	19 mths	Used for ECG and culled 26/02/2015, heart collected for histology
Rlf$^{MommeD8/MommeD34}$	G2.1.7.14	F	7.11.12	28.11.12	24 mths	Culled 26/11/14, eyes deteriorated and had growth on right ovary
Rlf$^{MommeD8/MommeD34}$	23.12.8	F	10.5.13	31.5.13	20 mths	Culled 08/01/15, one eye deteriorated and had growth on right ovary
Rlf$^{MommeD8/MommeD34}$	23.12.23	F	5.8.13	26.8.13	17 mths	Culled 08/01/15, no ovarian growth, right kidney had cyst
Rlf$^{MommeD8/MommeD34}$	23.12.25	F	5.8.13	26.8.13	17 mths	Culled 08/01/15, no abnormalities observed
Rlf$^{+/+}$	14	F	9.5.13	30.5.13	20 mths	Culled 08/01/15, for histology match
Rlf$^{+/+}$	16	F	9.5.13	30.5.13	20 mths	Culled 08/01/15, for histology match
Rlf$^{+/+}$	40	F	10.8.13	31.8.13	17 mths	Culled 08/01/15, for histology match
Rlf$^{+/+}$	41	F	10.8.13	31.8.13	17 mths	Culled 08/01/15, for histology match

From the females of this cohort two *Rlf*$^{MommeD8/MommeD34}$ mice (aged 24 and 20 months) were culled due to signs of physical distress, examination of each revealed they both had a fluid filled growth on their right ovary, **Figure 5.5** and data not shown. Following the observation of this phenotype in the second female the decision was made to cull the rest of the female cohort (aged 17 – 20 months). Of the two remaining *Rlf*$^{MommeD8/MommeD34}$ females no ovarian growth was observed in either mouse. Although a cyst on the right kidney of one female was observed (data not shown). None of the female wild-type mice displayed any gross morphological abnormalities. Furthermore, this phenotype has not been observed in larger numbers (n = 50) of aged wild-type females maintained by our laboratory (Dr S. Harten, *personal communication*).

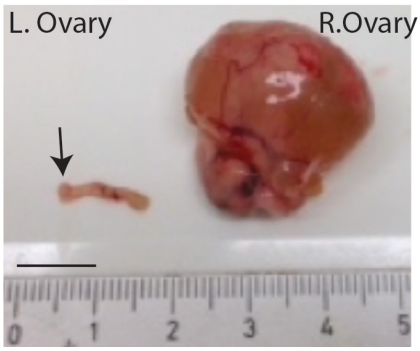


Figure 5.5: Ovarian growth observed in compound heterozygote female mutants.

Ovarian growth observed in *Rlf*$^{MommeD8/MommeD34}$ female right ovary. The left ovary and uterine horn from same mouse also included showing no growth present. Scale bar = 1 cm.

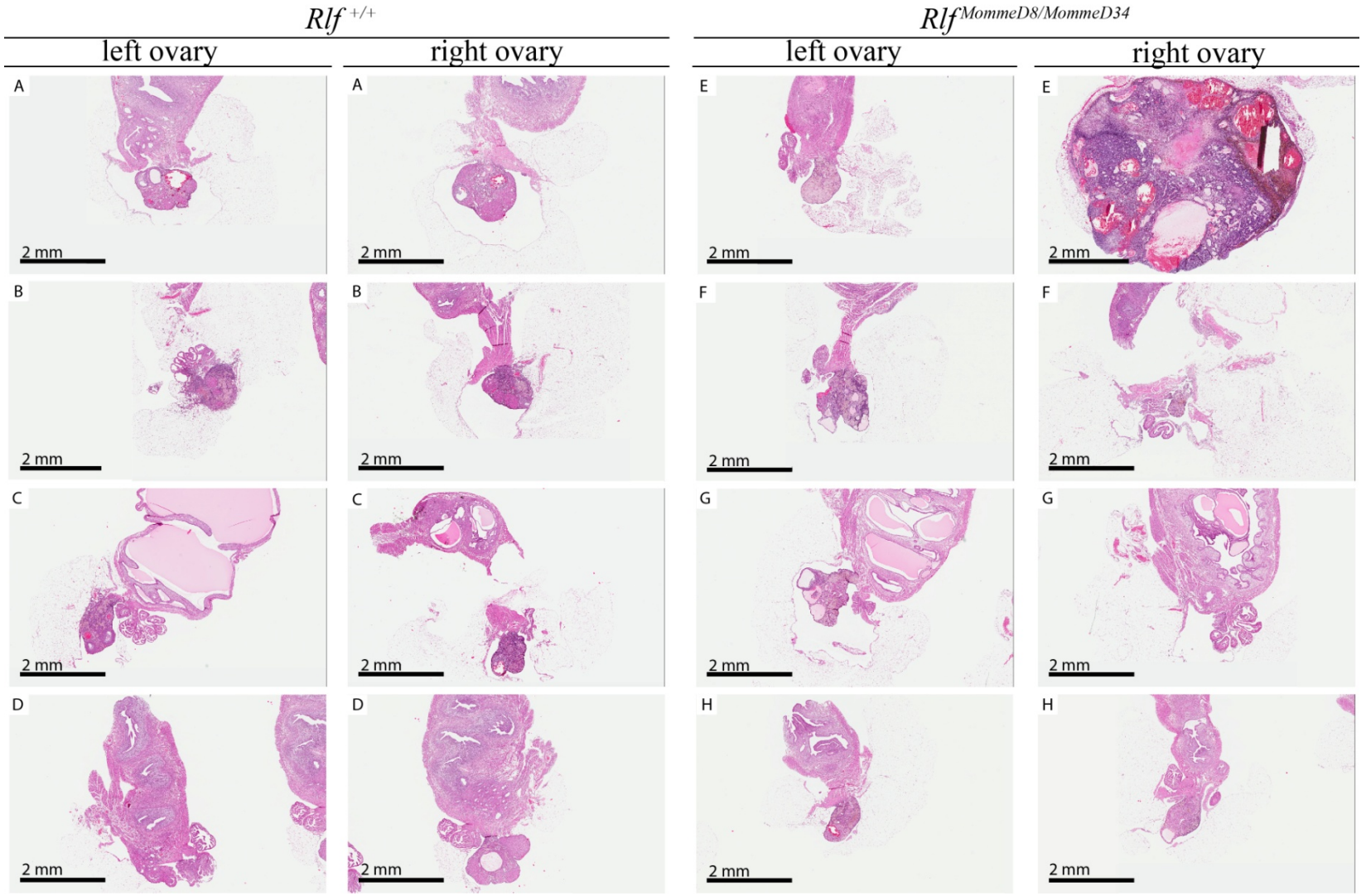
From each female I collected both the left and right ovaries and fixed them in 4% PFA for histological analysis, ovaries were then sectioned by QIMRB histopathology services**.** Slides were assessed in collaboration with Prof Dagmar Wilhelm, Monash University, an expert in gonad development and reproduction. Whilst ovarian growths were observed on two right ovaries of the four *Rlf* $^{MommeD8/MommeD34}$ females, histology of the remaining right and left ovaries showed differences in morphology compared to wild-type ovaries. In mammals, the ovary contains a number of different cell types including oocytes and somatic cells (granulose, theca and stoma cells), which act together to determine the development of follicles and ovulation (Richards & Pangas, 2010). No oocytes or follicles were observed in *Rlf* $^{MommeD8/MommeD34}$ ovaries (*Prof. D. Wilhelm, personal communication*); **Figure 5.6E-H** and Figure 5.7. The *Rlf* $^{MommeD8/MommeD34}$ right ovary with the fluid filled growth, **Figure 5.6E** and Figure 5.7, also appeared to have more stoma and granulose cells than normal, and were similar in appearance to follicular cysts (*Prof. D. Wilhelm, personal communication*). Due to time constraints in this PhD it was not feasible to generate a larger independent cohort of aged females to study this phenotype in greater detail. Additionally, it would also be important to determine whether this phenotype is present in younger females or arises from earlier in development, this will be discussed in **Section 5.3**.

Next Page:

Figure 5.6: Ovarian growth found in *Rlf* $^{MommeD8/MommeD34}$ aged females may be granulose cell tumours.

Haematoxylin and eosin staining of aged ovaries from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD8/MommeD34}$ females. Right ovaries E and F were observed to have large growths following dissection. *n* = min. 4 per genotype, scale bar = 2 mm.

Figure 5.6A: Ovarian growth found in *Rlf*$^{MommeD8/MommeD34}$ aged females may be granulose cell tumours.



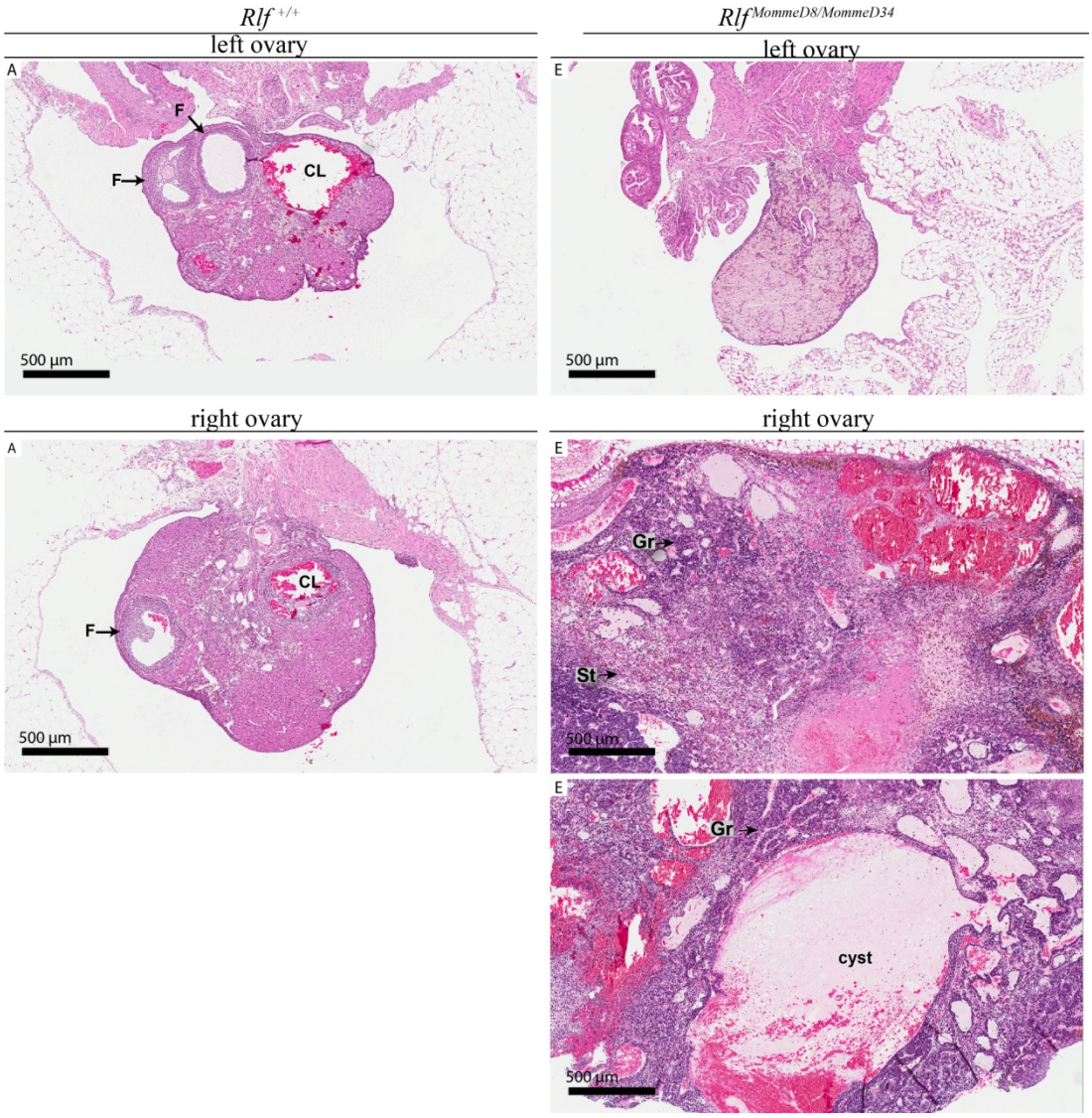


Figure 5.7: Higher magnification images of representative *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ aged ovaries.

Representative higher magnification images of wild-type (A) and *Rlf* $^{MommeD8/MommeD34}$ (E) (Figure 5.6), females highlighting follicles, and corpus lutea (F, CL) in wild-type mice not observed in mutant mice, and stromal (St), and granulosa (Gr) cells observed in *Rlf* $^{MommeD8/MommeD34}$ ovaries. Scale bar = 500 µm

5.2.6 Reduced dosage of *Rlf* has no impact on adult heart function

Following the observation of a heart defect in *Rlf* $^{MommeD28/MommeD28}$ and *Rlf*$^{MommeD34/MommeD34}$ embryos, I wanted to determine whether reduced dosage for *Rlf* impacts upon heart structure and function in adult mice. The remaining males from the aged mice cohort generated and studied in **Section 5.2.5** were used for analysis. Although this was a small number of mice (n = 7 *Rlf*$^{+/+}$, 3 *Rlf*$^{MommeD8/MommeD34}$), this study was proposed to gather preliminary information on adult mouse heart structure and function. Compound heterozygous mice were used for this study due to *MommeD28* and *MommeD34* homozygous mice dying shortly after birth. As there are no facilities for performing echocardiography (ECG) at QIMRB *Rlf* $^{MommeD8/MommeD34}$ and *Rlf* $^{+/+}$ males were transferred to the University of Queensland, and experimentation carried out in conjunction with Dr E. Porello, who is experienced in performing mouse ECGs. ECG is a non-invasive and popular method for evaluating changes in left ventricular (LV) structure and function in transgenic mice (Gardin et al., 1995; Montgomery et al., 2007; Tanaka et al., 1996).

The structural characteristics evaluated in these mice were interventricular septal thickness (IVS), LV internal diameter (LVID), and LV posterior wall thickness (LVPW). Each of these was measured at both systole (contraction) (IVSs, LVIDs, LVPWs) and diastole (relaxation) (IVSd, LVIDd, LVPWd). A trend of decreased LVID was observed at both systole and diastole in *Rlf* $^{MommeD8/MommeD34}$ versus wild-type mice, however this was not statistically significant, **Figure 5.8A**. This trend in decreased internal diameter could be the result of reduced bodyweight in *Rlf*$^{MommeD8/MommeD34}$ mice compared to *Rlf* $^{+/+}$ mice (**Figure 5.4B**), as body mass

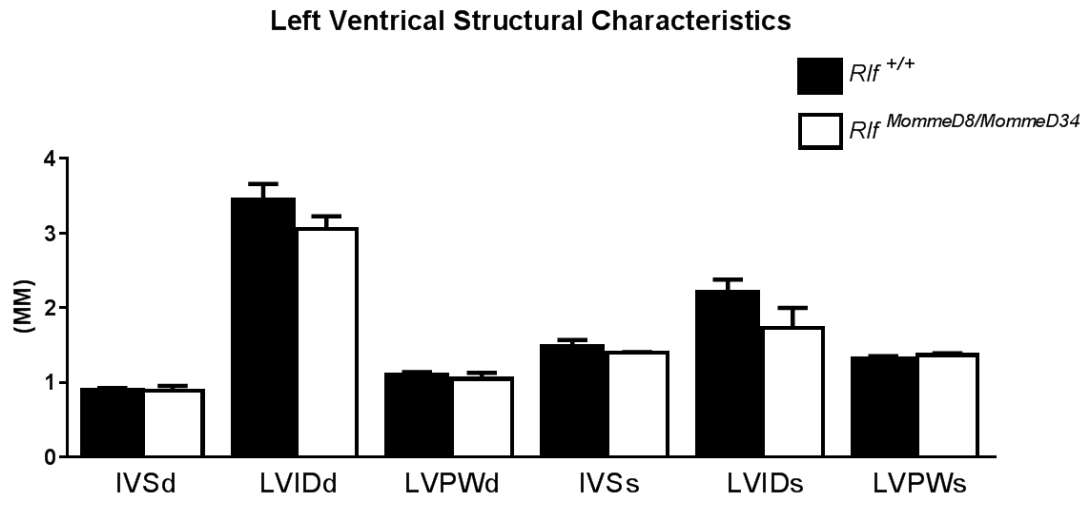
has been shown to correlate with LV mass (Gardin et al., 1995). In order to determine whether the trend in decreased LVID is a true effect, a minimum of 27 mice per genotype would need to be analysed (using a confidence level of 80% and 5% margin of error). Due to the reduced viability of *Rlf* $^{MommeD8/MommeD34}$ offspring, obtaining these numbers in an aged cohort may be unfeasible, thus undertaking this study in a younger cohort of mice may be a more appropriate study. This will be discussed further below. No difference was observed for the remaining structural measurements between the two genotypes, **Figure 5.8A**.

Ejection fraction (the percentage of blood pumped from the heart during systole) and fractional shortening (an estimate of myocardial contractility calculated from LV diameter at end systole and diastole), were the two functional characteristics investigated. A non significant increase in both ejection fraction and fractional shortening was observed in the *Rlf* $^{MommeD8/MommeD34}$ mice compared to wild-types, **Figure 5.8B**. Studies in humans have shown increases in LV dilation can lead to reductions in ejection fraction (Lewis & Sandler, 1971). This correlates with the difference observed in LV internal diameter and ejection fraction between genotypes in this study.

In order to confirm this is a true difference between genotypes, a minimum of 18 mice per genotype would need to be analysed (using a confidence level of 80% and 5% margin of error). Again, due to the reduced viability of *Rlf* $^{MommeD8/MommeD34}$ offspring, obtaining these numbers in an aged cohort may be unfeasible, thus undertaking this study in a younger cohort of mice may be a more appropriate study

design. For example, analysing late gestation homozygous null (*MommeD28* or *MommeD34*) embryos may provide more information on how loss of *Rlf* affects cardiac structure and function. Studies investigating cardiac function in pre-natal and neonatal mice are regularly reported in the current epigenetic/cardiovascular literature (Montgomery et al., 2007; Wang et al., 2004).

A.



B.

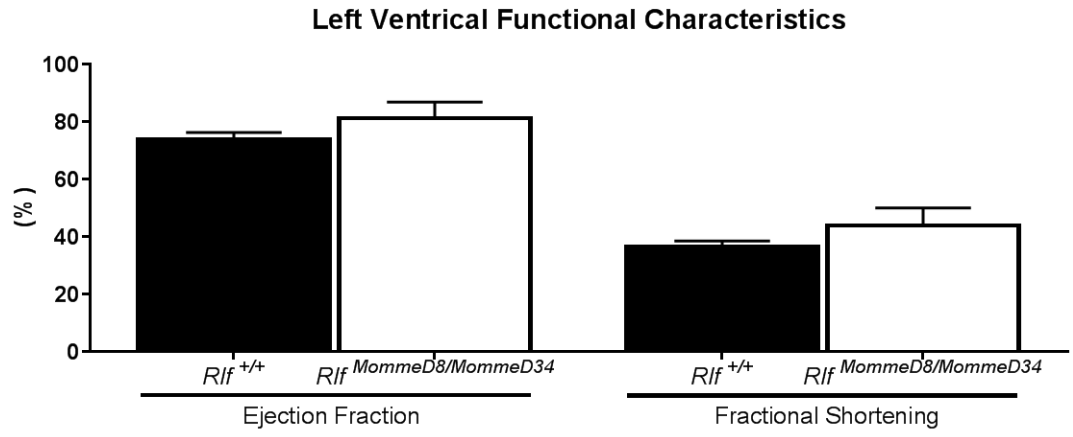


Figure 5.8: No differences in cardiac structure or function was observed in aged *Rlf*$^{MommeD8/MommeD34}$ mice.

A. Interventricular septal thickness (IVS), left ventricular internal diameter (LVID), and left ventricular posterior wall thickness (LVPW) at both diastole (IVSd, LVIDd, LVPWd) and systole (IVSs, LVIDs, LVPWs) were measured to analyse structural differences between *Rlf* $^{MommeD8/MommeD34}$ and *Rlf* $^{+/+}$ mice. A trend of decreased LVIDd and LVIDs was observed in compound heterozygous mice, but this was not statistically significant. B. No significant difference in ejection fraction or fractional shortening was observed *Rlf* $^{MommeD8/MommeD34}$ and *Rlf* $^{+/+}$ mice. Error bars represent SEM and Student's t-test was used to calculate p values, n = 7 *Rlf* $^{+/+}$ and 3 *Rlf* $^{MommeD8/MommeD34}$ mice.

5.2.7 Presence of the *MommeD8* allele alters methylation at endogenous loci

A second *Rlf* hypomorphic line with low Rlf protein activity is the *MommeD8* mouse line. Similar to hypomorphic *Rlf* $^{MommeD8/MommeD34}$ mice, *Rlf* $^{MommeD8/MommeD8}$ mice display reduced viability and weight that is not as severe as that observed in *Rlf* null mice (Ashe et al., 2008). Previous research has also found methylation of the GFP transgene is increased in *Rlf* $^{MommeD8/MommeD8}$ offspring (Daxinger et al., 2013), making these mice ideal for studying the effects of reduced *Rlf* expression on endogenous methylation.

To address this I undertook bisulphite sequencing analysis, of the *Smad3/Smad6* loci previously identified as being differentially methylated in both *Rlf* $^{MommeD28/MommeD28}$ and *Rlf* $^{MommeD34/MommeD34}$ E14.5 liver (**Section 3.2.5**). DNA from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD8/MommeD8}$ littermates (n = 2 per genotype) was bisulphite converted and sequenced as described previously (**Section 3.2.4**). In earlier studies I observed that in *Rlf* $^{MommeD28/MommeD28}$ fetal liver a 26% increase in methylation at the *Smad3/Smad6* region was observed, and a 40% increase observed in *Rlf* $^{MommeD34/MommeD34}$ fetal liver compared to wild-type littermates (**Figure 3.11**) (Harten et al., 2015). Bisulphite sequencing of fetal liver from the *MommeD8* mouse line found a ~14% increase in methylation, p <0.0014, in *Rlf* $^{MommeD8/MommeD8}$ embryos compared to *Rlf* $^{+/+}$ littermates, **Figure 5.9.** These findings suggest that carrying two copies of the hypomorphic *MommeD8* allele impacts upon methylation at endogenous loci, but the effect is not as large as in offspring carrying two null *Rlf* alleles, and this correlates with previous findings investigating endogenous gene expression (**Figure 5.2**).

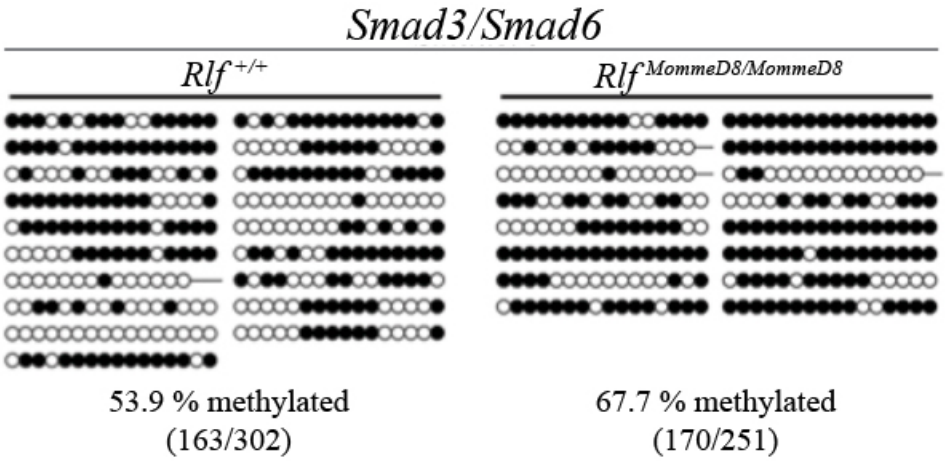


Figure 5.9: Methylation analysis of the *Smad3/Smad6* region in *Rlf*$^{+/+}$ and *Rlf*$^{MommeD8/MommeD8}$ fetal liver.

The *MommeD8* mutant allele results in an increase in methylation at a Rlf-DMR between *Smad3* and *Smad6* on Chromosome 9. Each column represents DNA from a single embryo (n = 2 for each genotype), each row is the sequence from a single cell and each circle represents one CpG site. Black circles represent methylated CpGs, white circles represent unmethylated CpGs, and lines (-) represents a non CpG or ambiguously sequenced position where a CpG exists in the genomic sequence. (X^2, p <0.05).

5.2.8 *Rlf*$^{MommeD8/MommeD8}$ dams produce smaller litters that weigh less than litters born from *Rlf*$^{MommeD8/MommeD8}$ sires

To assess the fertility and viability of *MommeD8* mutants I crossed three *Rlf*$^{MommeD8/MommeD8}$ sires to three wild-type dams and also performed the reciprocal cross with three *Rlf*$^{MommeD8/MommeD8}$ dams and wild-type sires. Mice were checked for the presence of a vaginal plug to determine mating had occurred and monitored for litters as a result of mating over 30 weeks. All offspring produced were heterozygous for the *MommeD8* allele.

The number of pups in each litter produced from crossing a *Rlf*$^{MommeD8/MommeD8}$ dam to a wild-type sire was found to be significantly smaller at both birth (average litter = 6.9 pups) and weaning (average litter = 5.4 pups) when compared to litters produced from a *Rlf*$^{MommeD8/MommeD8}$ sire mating at birth (9.3 pups) and wean (7.8 pups), **Figure 5.10A**. Comparison of numbers at birth versus weaning found no significant difference from either reciprocal cross, **Figure 5.10A**. A significant difference at weaning in litter size from *Rlf* $^{MommeD8/MommeD8}$ dams (5.38 pups), but not *Rlf*$^{MommeD8/MommeD8}$ sires (7.8 pups), compared to wild-type controls (7.30 pups) was also observed, **Figure 5.10B.** Slightly more females than males were observed at weaning from the *Rlf* $^{MommeD8/MommeD8}$ sire cross (n = 47 females, 39 males), and similar numbers of males to females were observed from the *Rlf*$^{MommeD8/MommeD8}$ dam cross (n = 48 and 49 respectively), **Figure 5.10C.**

No difference in survival rates of pups born from small versus large litters, or males versus females, from either cross was observed. With the same numbers in pup

death observed between the respective groups (data not shown). This suggests the differences observed between the reciprocal crosses may be due to differences in maternal fertility as the number of pups per litter is significantly reduced from birth.

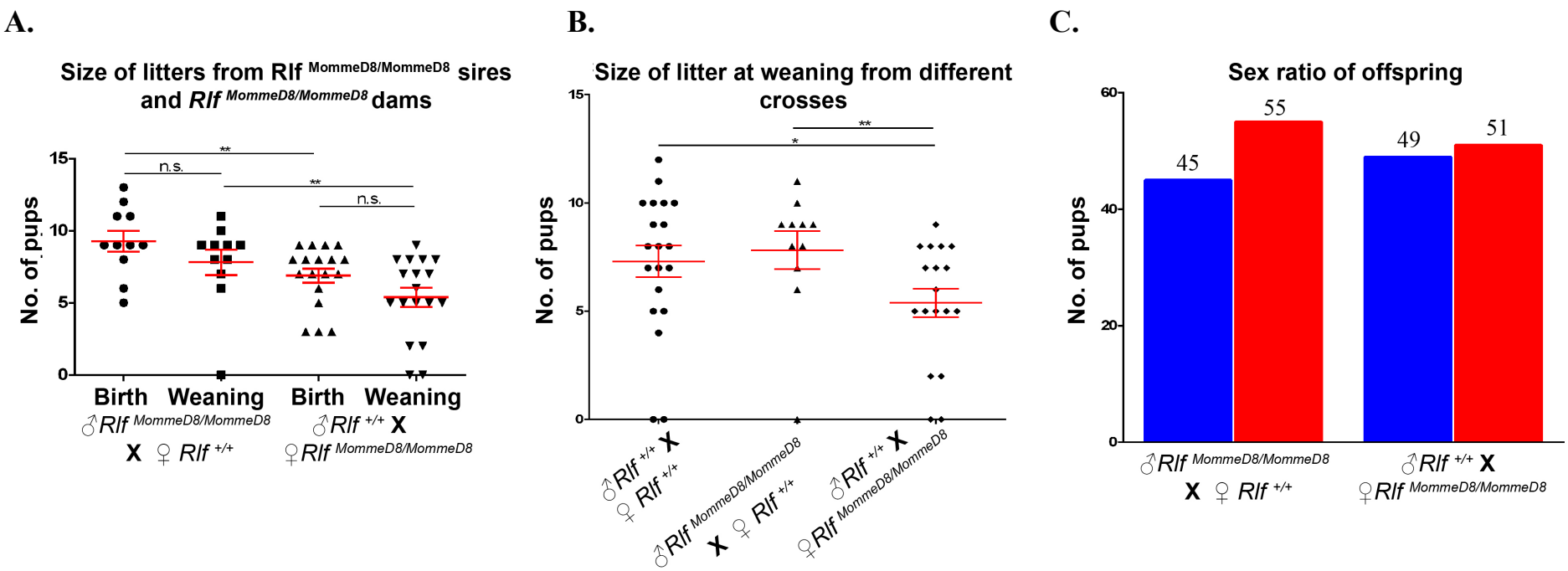


Figure 5.10: *Rlf*$^{MommeD8/MommeD8}$ dams produce smaller litters than litters born from *Rlf*$^{MommeD8/MommeD8}$ sires or wild-type parents.

A. Litters from crosses between ♀*Rlf*$^{MommeD8/MommeD8}$ and ♂*Rlf*$^{+/+}$ mice have reduced number of offspring, when compared to the reciprocal cross of ♂*Rlf*$^{MommeD8/MommeD8}$ and ♀ *Rlf*$^{+/+}$ (n =11 and 18 litters respectively). The difference in litter size is observed from birth. Each data point represents an individual pup. Error bars represent SEM and Mann-Whitney U-test (A,B) or X^2 test (C) was used to calculate p values (* p <0.05, ** p < 0.01). **B.** Weaning data from (A) showing litters born from *Rlf*$^{MommeD8/MommeD8}$ dams are smaller than those from wild-type parents and *Rlf*$^{+/+}$ sires. (n = 18 (wild-type), 11 (sire) and 18 (dam) litters). **C.** Similar ratios of male (blue, n = 86) to female (red, n = 97) offspring were observed when comparing crosses undertaken between ♀*Rlf*$^{MommeD8/MommeD8}$ and ♂ *Rlf*$^{+/+}$ and ♂ *Rlf*$^{MommeD8/MommeD8}$ x ♀ *Rlf*$^{+/+}$ mice.

Following the observation of differences in litter sizes produced from *Rlf*$^{MommeD8/MommeD8}$ dams and sires, I wanted to also investigate whether there was any maternal or paternal impact on weight. The weights of all offspring was measured at weaning (21 days), and offspring conceived from *Rlf*$^{MommeD8/MommeD8}$ dams were found to weigh significantly less (9.88 g, ± 0.254) than offspring conceived from *Rlf*$^{MommeD8/MommeD8}$ sires (10.73 g, ± 0.166), **Figure 5.11A**. Females produced from *Rlf*$^{MommeD8/MommeD8}$ dams weighed significantly less (9.43 g, ± 0.341) than females conceived from *Rlf*$^{MommeD8/MommeD8}$ sires (10.73 g, ± 0.246), but did not weigh less than their male littermates, **Figure 5.11A**.

When bodyweight was normalised for each litter so as to compare the weight of pups to the average weight of the litter in which they were born. Only female offspring born from a *Rlf*$^{MommeD8/MommeD8}$ sire were found to have reduced weight compared to their male littermates, but this did not impact upon the overall weight of all offspring, **Figure 5.11B**. No difference was observed when comparing male and female offspring born from *Rlf*$^{MommeD8/MommeD8}$ dams or comparing sexes across the reciprocal crosses, **Figure 5.11B.** Large variability in pup weight was observed across litters born from both *Rlf*$^{MommeD8/MommeD8}$ dams and *Rlf*$^{MommeD8/MommeD8}$ sires, suggesting more factors may be responsible for the fluctuations in weights observed, including competition for food amongst littermates which would differ based on litter size.

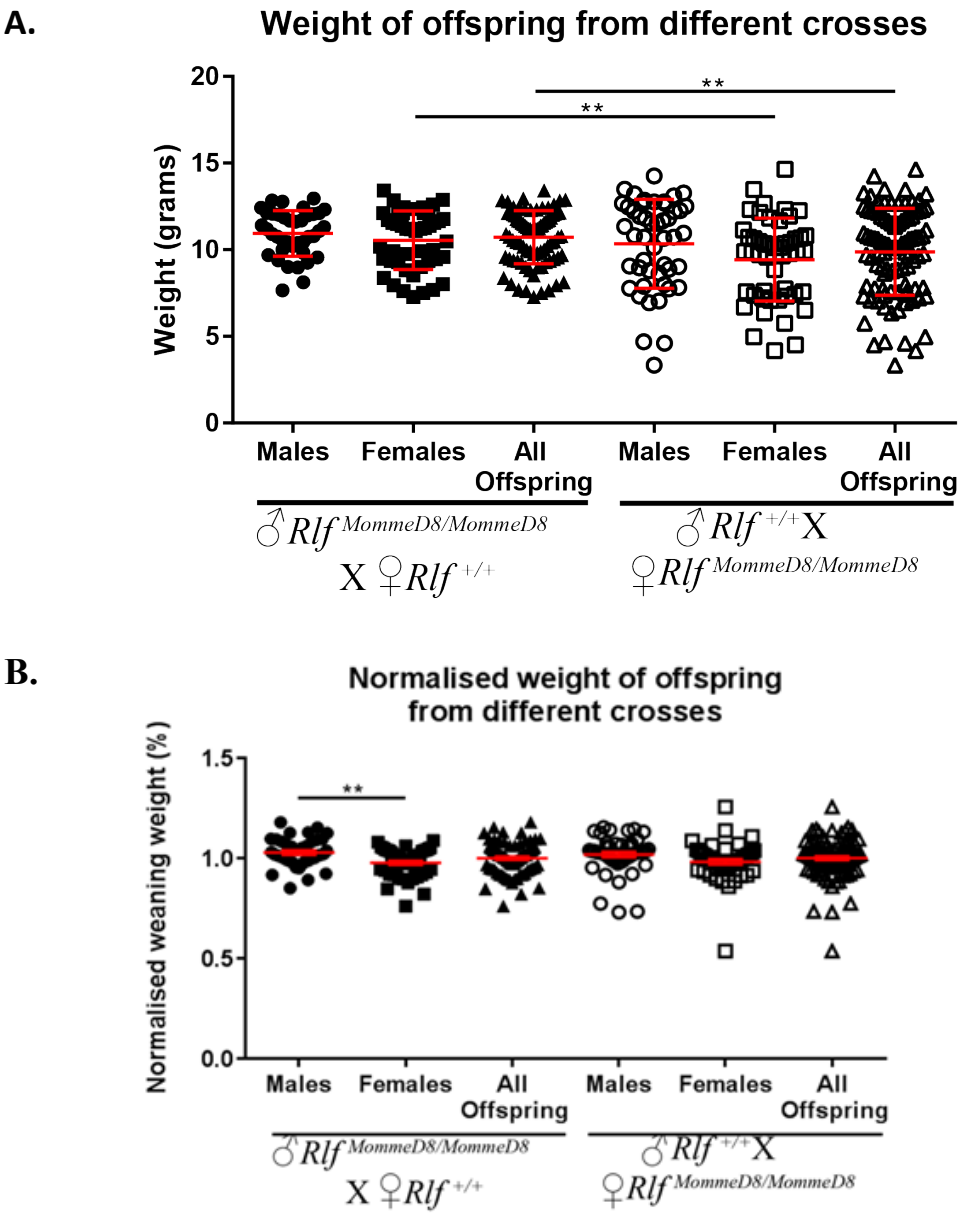


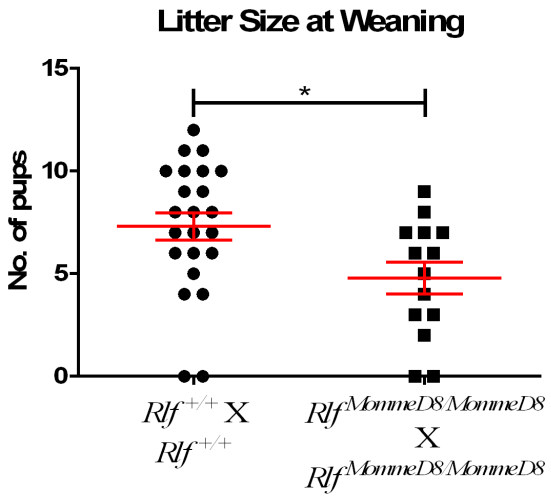
Figure 5.11: *Rlf*$^{MommeD8/MommeD8}$ dams produce litters that weigh less than litters born from *Rlf*$^{MommeD8/MommeD8}$ sires.

A. Bodyweights of offspring from different crosses reveals a reduction in bodyweight in offspring born from a *Rlf* $^{MommeD8/MommeD8}$dams (n = 97) cross compared to those born from a *Rlf* $^{MommeD8/MommeD8}$ sires (n = 86). B. The weight of offspring were normalized to the average weight of the litter in which they were born. Only females born from a *Rlf* $^{MommeD8/MommeD8}$ sire were found to be smaller on average versus their male littermates. Each data point represents an individual pup. Error bars represent SEM and Student's t-test was used to calculate p values (** $p <$ 0.01).

5.2.9 Effect of reduced *Rlf* expression on weight across generations

To investigate the effect of reduced *Rlf* expression across generations, matings of *Rlf* $^{MommeD8/MommeD8}$ mice were established to produce a transgenerational cohort. F1 offspring were produced from a *Rlf* $^{MommeD8/+}$ intercross resulting in wild-type, heterozygous and homozygous offspring. F1 litter sizes were found to be significantly smaller than litters born from two wild-type parents, **Figure 5.12A**. This was consistent with previous findings from Ashe et al., 2008. These authors also found *MommeD8* homozygotes from *Rlf* $^{MommeD8/+}$ intercrosses weigh significantly less than both their wild-type and heterozygous littermates at weaning (Ashe et al., 2008). In this study I found F1 *MommeD8* homozygous offspring to weigh significantly less than heterozygous littermates but not wild-type littermates, **Figure 5.12B.** While the difference in weight between wild-type and homozygous *MommeD8* offspring was not significant, a trend towards decreased weight in homozygotes was observed, reflecting the results found by Ashe et al., 2008. Also of note is that the mean weights of mice in this study were larger (*Rlf* $^{+/+}$ 10.54g; *Rlf* $^{MommeD8/+}$ 11.26g; *Rlf* $^{MommeD8/MommeD8}$ 9.41g) than those previously published (*Rlf* $^{+/+}$ 8.54g; *Rlf* $^{MommeD8/+}$ 8.65g; *Rlf* $^{MommeD8/MommeD8}$ 6.60g), but did not conflict with previous results.

A.



B.

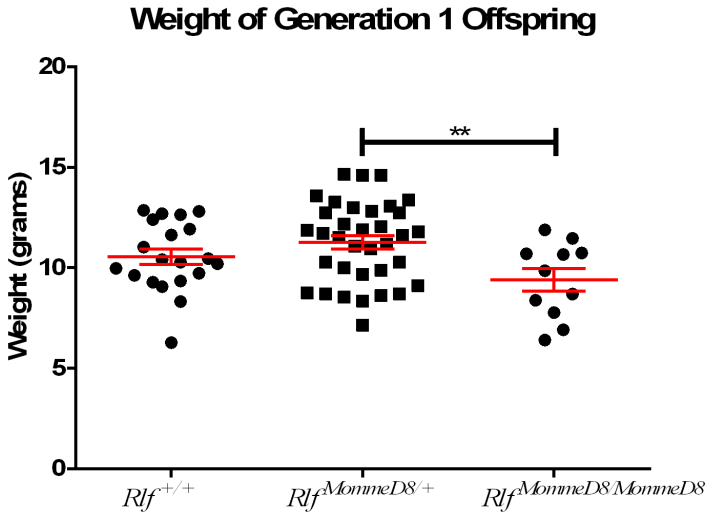


Figure 5.12: *Rlf*$^{MommeD8/+}$ intercross F1 litters are smaller than *Rlf*$^{+/+}$ litters, and homozygous offspring weigh less.

A. The number of pups in each litter born from a *Rlf* $^{MommeD8/MommeD8}$ intercross is reduced to the number of pups born from a *Rlf* $^{+/+}$ intercross (n = 14 and 21 litters respectively). Each data point represents an independent litter B. Homozygous offspring from a *Rlf* $^{MommeD8/MommeD8}$ intercross are smaller at weaning than *Rlf* $^{MommeD8/+}$ littermates but not wild-type littermates. Each data point represents an individual mouse. Error bars represent SEM and Mann-Whitney U-test was used to calculate p values (* $p < 0.05$, ** $p < 0.01$).

Homozygous *MommeD8* F1 offspring were then mated together to produce F2 offspring, where all offspring were *MommeD8* homozygotes. F2 offspring were then mated together to produce F3 offspring and this mating was continued for 8 generations. Analysis of litter size at each generation compared to wild-type litter size found a significant reduction in the number of pups in each generation, except for F5, versus wild-type, **Figure 5.13A.** A significant decrease in litter size was also observed in F6 versus F5, **Figure 5.13A.** This is likely due to a reduction in the number of F6 litters surviving to weaning. No significant difference in litter size was observed comparing each of the other homozygous generations to one another.

The weights of offspring across generations was also analysed. Significant reductions in weights were observed in F4 versus F3 offspring and F7 versus F6 offspring, **Figure 5.13B**. In contrast, significant increases in weights were observed in F6 offspring versus F5, and F8 offspring versus F7, **Figure 5.13B**. Interestingly, the increased weight of F6 pups at weaning versus F5 correlates with the decrease in F6 litter size observed, **Figure 5.13A.** No statistical differences in litter sizes were observed between the other generations in which statistical differences in mouse weight were observed. A trend towards increased weight in mice from smaller litters and decreased weights in mice from large litters was also observed. To determine the correlation between litter size and weight analysis was undertaken on data collected across all of the generations. A negative correlation ($R^2 = 0.09237$, $p < 0.0001$) between litter size and weight was observed, **Figure 5.14**, and these results were comparable to independent studies investigating the same parameters (Chahoud & Paumgartten, 2009; Johnson et al., 2001). The absence of a cumulative effect

towards increased or decreased weights suggests it is unlikely reduced *Rlf* expression is responsible for the differences observed.

Next Page:

Figure 5.13: No cumulative effects on litter size or offspring weight is observed across *MommeD8* generations.

A. Litter size at weaning reveals a reduction in size in F1-8 litters versus wild-type litters, and F6 litters versus F5. Each data point represents an independent litter. B. Weight of offspring at weaning reveals a correlation between litter size (A) and offspring weight (B), whereby offspring from smaller litters are likely to have larger bodyweights than those from large litters. Each data point represents an individual mouse. Error bars represent SEM and Mann-Whitney *U*-test was used to calculate *p* values (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$).

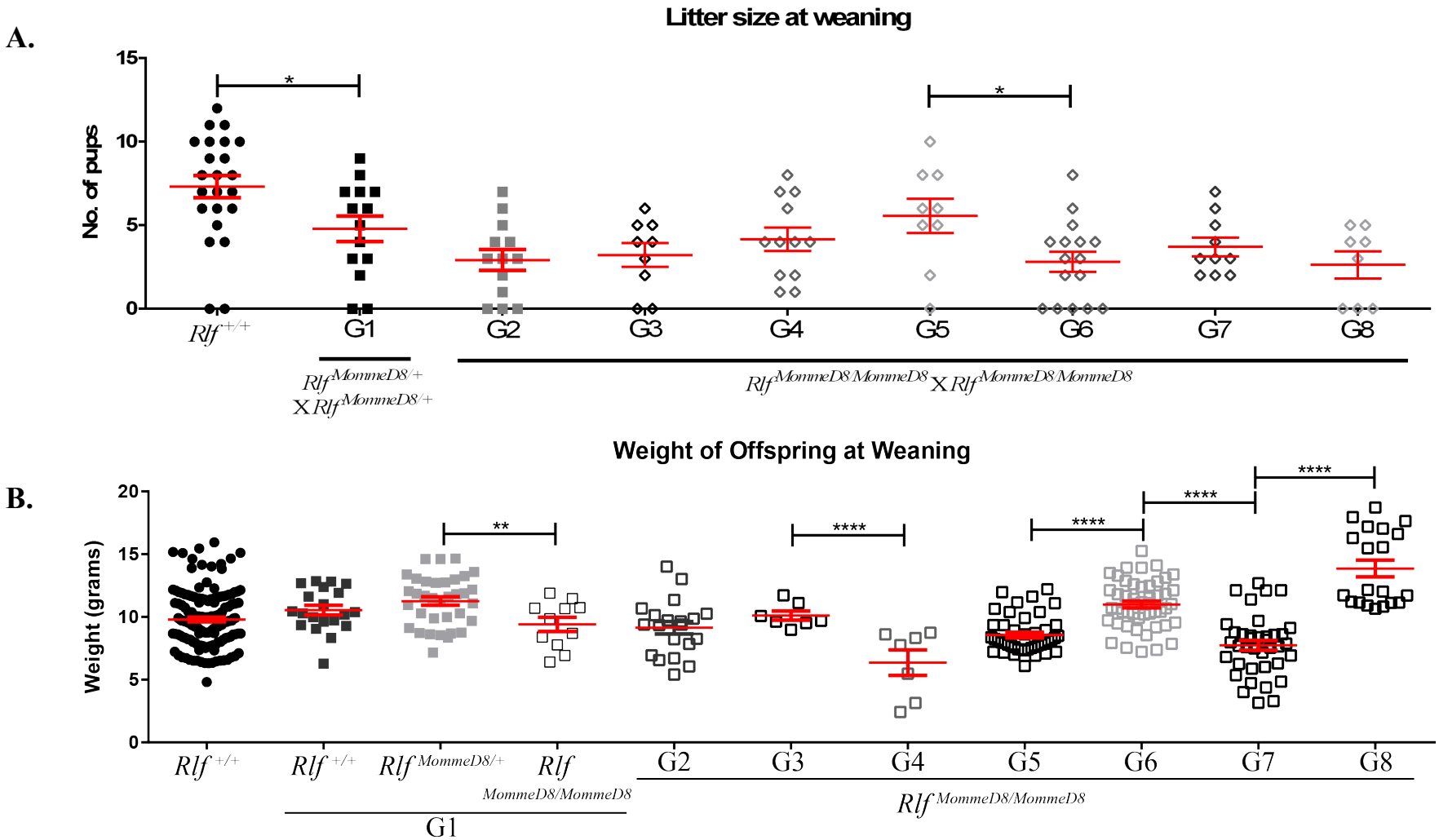


Figure 5.13: No cumulative effects on litter size or offspring weight is observed across *MommeD8* generations.

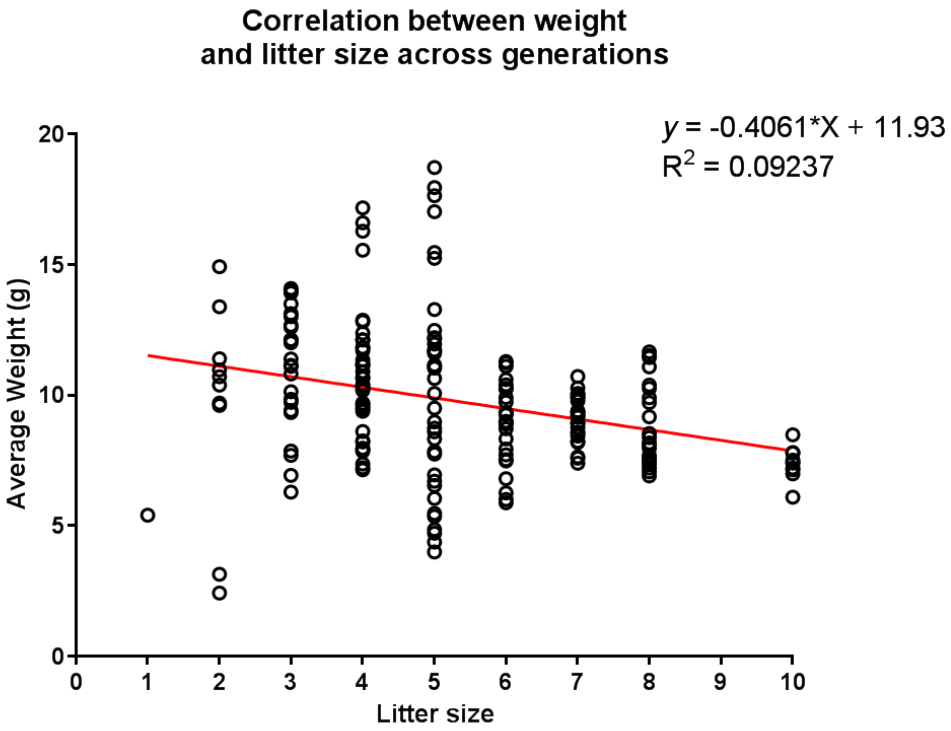


Figure 5.14: A slight correlation between litter size and pup weight is observed across all *MommeD8* generations.

Relationship between litter size and pup weight at wean for all litters born across eight *MommeD8* generations, shows a slight correlation between the two (R^2 = 0.09237) (n = 46 litters, 207 pups). Line represents linear regression slope as calculated by GraphPad Prism 6.

5.2.10 Effect of reduced *Rlf* dosage on methylation at endogenous loci across generations

Reduced *Rlf* expression has previously been shown to alter methylation of the GFP transgene in *Rlf* $^{MommeD8/MommeD8}$ mice compared to *Rlf* $^{+/+}$ mice (Daxinger et al., 2013). As a result of this previous finding I chose to investigate methylation of the GFP transgene in early and late generation (F1 and F8) adult liver. Tissue was collected from 3 week old mice from each generation (n = 2 per generation), as well as wild-type controls (n = 2), and underwent bisulphite conversion and sequencing as described previously (**Section 3.2.4**). A small increase in methylation in F1 *Rlf* $^{MommeD8/MommeD8}$ was observed compared to wild-type controls (69% and 63% respectively), however this was not significant, p 0.0996, **Figure 5.15**. A further small increase in F8 offspring (73%) was observed compared to F1 offspring (69%), but this was also not statistically significant, $p < 0.0931$, **Figure 5.15**. Whilst these increases are small, carrying out breeding for a greater number of generations may reveal that this methylation changes increases slowly but steadily across the generations.

Following the observation of a small difference in methylation between generations at the GFP transgene, I wanted to investigate the impact of reduced *Rlf* expression across generations on methylation of an endogenous region. The *Smad3/Smad6* region, found to be differentially methylated in fetal liver in each of the *Rlf Momme* lines (**Figure 3.11** and **Figure 5.9**), was chosen for further investigation. Methylation of this region in *Rlf* $^{+/+}$ adult liver was observed to be much higher (81%) than in *Rlf* $^{+/+}$ fetal liver (53%), **Figure 5.15** and **Figure 5.9**.

Again, small increases in methylation in F1 and F8 *Rlf* $^{MommeD8/MommeD8}$ offspring at this region were observed compared to wild-types (F1- 84%, p=0.4831; F8- 88%, p=0.0152; and wild-type 81%), **Figure 5.15**. However, the difference in methylation between F1 mice and wild-types (3%, p=0.4831) was not as great as that observed in *Rlf* $^{MommeD8/MommeD8}$ fetal liver compared to wild-types (14%, p=0.0014, **Figure 5.9**). As such investigating methylation of this region in adult liver may not have been an appropriate tissue choice. Repeating the experiment using fetal liver tissue across the generations may reveal the small changes observed in adult liver to be more pronounced in fetal liver tissue, helping to define whether reduced *Rlf* expression has an effect on epigenetic inheritance across generations. This will be discussed further in **Section 5.3.**

Next Page:
Figure 5.15: Methylation is not altered in late generation *MommeD8* homozygotes.

Bisulphite sequencing of the GFP transgene (**A**) and *Smad3/Smad6* region (**B**) reveals subtle differences in DNA methylation in *Rlf* $^{MommeD8/MommeD8}$ mice compared to wild-types (n = 2 per genotype). Each column represents DNA from a single heart, each row is the sequence from a single cell and each circle represents one CpG site. Filled circles represent methylated CpGs. Black circles represent methylated CpGs, white circles represent unmethylated CpGs, and lines (-) represents a non CpG or ambiguously sequenced position where a CpG exists in the genomic sequence. (X^2, GFP transgene: p=0.0996 WT v F1, p=0.0931 WT v F8, p=0.9259 F1 v F8; *Smad3*/*Smad6*: X^2, p=0.4831 WT v F1, p=0.0152 WT v F8, p=0.2523 F1 v F8).

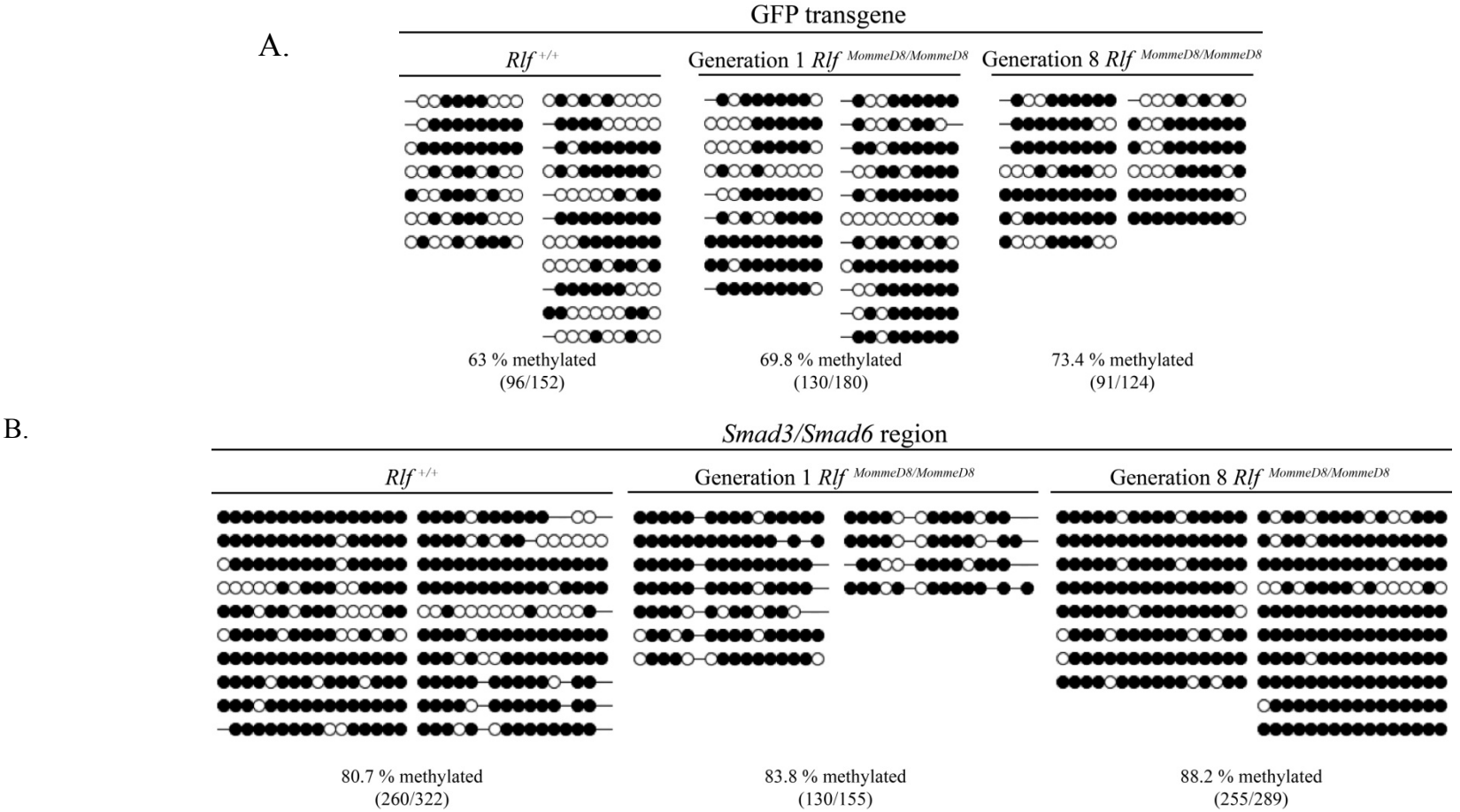


Figure 5.15: Methylation is not altered in late generation *MommeD8* homozygous offspring.

5.3 DISCUSSION

Both *MommeD28* and *MommeD34* homozygous offspring are non-viable at weaning, whereas *MommeD8* homozygous offspring are viable but present in less than expected numbers (**Table 4.1** and Daxinger et al, 2013). Each of the *Rlf Momme* lines have also been shown to alter expression and methylation of the GFP transgene, which was designed to be sensitive to epigenetic silencing (Ashe et al., 2008; Daxinger et al., 2013). In this chapter I have further investigated the effect of reduced *Rlf* in *Rlf* $^{MommeD8/MommeD8}$ mice on endogenous gene expression and methylation. Most of the loci previously found to be either differentially expressed or methylated in *Rlf* $^{MommeD28/MommeD28}$ or *Rlf* $^{MommeD34/MommeD34}$ mice were also found to be significantly differentially expressed (**Figure 5.1**) or methylated (**Figure 5.9**) in *Rlf* $^{MommeD8/MommeD8}$ mice. Furthermore, the presence of two *MommeD8* alleles had a greater impact on gene expression than that observed in *MommeD34* heterozygous mice, but less than that observed in *MommeD34* homozygous mice (**Figure 5.2**). This data supports the previous findings investigating viability of the *Rlf Momme* mouse lines; that carrying two copies of the *MommeD8* allele has a greater impact on survival and expression than carrying one copy of the *MommeD34* allele in the mouse (Ashe et al., 2008; Daxinger et al., 2013).

The previous studies investigating the effect of mutations in *Rlf* on viability in the mouse investigated the different *Rlf Momme* lines independently from each other. Here I undertook breeding of *Rlf* $^{MommeD8/MommeD8}$ mice with *Rlf* $^{MommeD34/+}$ mice to investigate the viability of compound heterozygous (*Rlf* $^{MommeD8/MommeD34}$) mice. The presence of both the *MommeD8* and *MommeD34* alleles in mice was found to reduce

viability and also weight of *Rlf* $^{MommeD8/MommeD34}$ offspring at weaning (**Figure 5.4**). No difference in Rlf expression was observed in *Rlf* $^{MommeD8/MommeD34}$ mice, however *Hpd* mRNA expression was significantly altered in these mice compared to wild-type and heterozygous *MommeD28* and *MommeD8* mice (Figure 5.3). The development of compound heterozygous/ hypomorphic mouse lines are beneficial for studying the biological impact of mutations in cases where null expression results in lethality. The availability of the different *Rlf* mutant mouse lines provides a unique tool for studying the effects of reduced *Rlf* expression on phenotype that may reflect occurrences in the human population.

Unlike the null *Rlf* mouse lines, *MommeD28* and *MommeD34,* some compound heterozygous mice are viable at three weeks of age (**Figure 5.4**). Following the observation of differences in cardiac morphology in homozygous null *Rlf* embryos, the viability of *Rlf* $^{MommeD8/MommeD34}$ mice provided an opportunity to determine whether reduced *Rlf* expression alters cardiac structure and function in the adult heart. No difference in either cardiac structure or function was observed between *Rlf* $^{+/+}$ and *Rlf* $^{MommeD8/MommeD34}$ adult mice (**Figure 5.8).** This lack of difference in structure and function alongside of reduced survival of *Rlf* $^{MommeD8/MommeD34}$ offspring to adulthood may reflect incomplete phenotypic penetrance in compound heterozygous mice. Potentially, the *Rlf* $^{MommeD8/MommeD34}$ offspring that do not survive to adulthood may present with a more severe cardiac phenotype that is reminiscent of the phenotype observed in the null *Rlf* lines, compared to offspring that do survive to adulthood. Undertaking analysis of cardiac function and structure in late stage embryos may allow us to more accurately define whether compound heterozygous embryos display different severities in cardiac phenotype and in turn enable the

detection of which offspring will survive to adulthood and which offspring may require the implementation of intervention strategies for survival.

Two of the *Rlf* $^{MommeD8/MommeD34}$ aged females were found to have ovarian growths (**Figure 5.5**), and while no growths were observed on other *Rlf* $^{MommeD8/MommeD34}$ ovaries, histological analysis found all of the ovaries examined had no oocytes or follicles present compared to wild-type ovaries (**Figure 5.6**). The reduced expression of *Rlf* in these mice may lead to impaired follicle development, resulting in the phenotype observed. Oocytes arise from primordial germ cells (PGCs), precursors of the postnatal female (and male) germ cell lineages, and become surrounded by somatic cells, forming primordial follicles (Ginsburg et al., 1990). The pool of primordial follicles formed just after birth makeup the entire supply of ovarian follicles during a mouse's reproductive life, thus reduced expression of *Rlf* early during development may impact upon the development and/or growth of primordial follicles. Investigating factors that are important for follicle growth, such as oestradiol (required for primordial follicle formation) and new ovary homeobox (required for primordial follicle growth), may define a specific time-point in which reduced *Rlf* is influencing follicle development.

The dysregulation of the passage of PGCs to oocytes can also influence tumorigenic transformation (Kraggerud et al., 2013). Many studies investigating ovarian tumorigenesis have utilised mice on the FVB/NJ background, of which the *Momme* mice are also on. FVB/NJ mice have been shown to be more susceptible to benign teratomas and granulosa cell tumours than other mouse strains (Mahler et al., 1996). This could mean that the *Momme* mice are more susceptible to defects in

ovarian development than other mouse strains, and a combination of factors could be influencing the ovarian phenotype observed in *Rlf* mutants (Hsu et al., 1996; Peltoketo et al., 2010; Youngson et al., 2011). Although the absence of changes in morphology in the wild-type aged mice from this cohort suggest the phenotype is due to reduced *Rlf* expression in *Rlf*$^{MommeD8/MommeD34}$ mice.

During this study litters born to *Rlf*$^{MommeD8/MommeD8}$ dams were found to contain fewer offspring than those born to *Rlf* $^{MommeD8/MommeD8}$ sires or wild-type dams, (**Figure 5.10**). Whilst ovarian morphology in the *MommeD8* line has not previously been investigated, the differences observed in *Rlf* $^{MommeD8/MommeD34}$ females may be linked to the reduction in litter size observed. Studies in ageing female mice have shown small litter size coincides with the degeneration of oocytes (Castrillon et al., 2003; Danilovich & Sairam, 2002). These studies have analysed serum levels of follicle stimulating hormone and luteinizing hormone, with increased levels of these hormones being linked to accelerated oocyte loss and follicle degeneration respectively (Danilovich & Sairam, 2002; Kumar & Sait, 2011). As *Rlf*$^{MommeD8/MommeD34}$ ovaries were found to contain no oocytes or follicles, an increase in the serum levels of these hormones may explain the histological observations. Alternatively, counting the number of oocytes released by compound heterozygous mice, by inducing superovulation, and may reveal whether the reduced litter sizes observed are due to *Rlf* $^{MommeD8/MommeD8}$ dams having fewer eggs available for fertilisation.

Intrauterine growth restriction (IUGR) may also be a contributing factor to the reduced litter sizes observed from *Rlf*$^{MommeD8/MommeD8}$ dams. IUGR is classified as a

lack of nutrient, oxygen and blood supply from the mother to the fetus (Corstius et al., 2005). Whilst we have not directly investigated the placentas of mice from the *MommeD8* mouse line, we have previously noted subtle differences in placental morphology of *Rlf*$^{MommeD28/MommeD28}$ embryos (**Section 4.2.8**). As such, it is possible that the reduction in litter size may be the result of differences in placental morphology affecting nutrient supply to embryos, resulting in a reduced number of pups at birth.

In contrast to homozygous *MommeD28* and *MommeD34* mice the majority of homozygous *MommeD8* mice are viable and fertile, presenting an opportunity to examine the effects of reduced *Rlf* expression across multiple generations. *Rlf*$^{MommeD8/MommeD8}$ mice have also been found to have decreased bodyweight (**Figure 5.12**) and increased methylation of the GFP transgene (Daxinger et al., 2013) compared to wild-type littermates. To determine whether reduced *Rlf* can influence epigenetic reprogramming in the mouse, the weight of offspring across multiple generations and also methylation of the GFP transgene and an endogenous region were investigated. No large accumulation of either of these phenotypes across the *MommeD8* generations was observed. Instead, the maintenance of bodyweight (**Figure 5.13**) and GFP methylation (**Figure 5.15**) in late generation *Rlf* $^{MommeD8/MommeD8}$ mice support the resetting of epigenetic state between each generation. The study of bodyweight across generations was found to be highly variable across generations (**Figure 5.13**) and may have been influenced by external factors, such as intermittent building works in the animal house at the time of the experiment, which could not be controlled for.

Methylation of the endogenous *Smad3/Smad6* locus was also investigated, and a small difference in methylation observed comparing wild-type, F1 and F8 *Rlf* $^{MommeD8/MommeD8}$ mice (**Figure 5.15**). The *Smad3/Smad6* region and GFP transgene were originally identified as being differentially methylated in E14.5 liver and adult spleen (**Figure 5.9** and Daxinger et al., 2013). As the methylome has been shown to undergo remodelling during development and to also be different between different lineages (Hemberger et al., 2009; Okano et al., 1999), the investigation of adult liver may not have been an appropriate tissue choice. The collection of either fetal liver and/or adult spleen for methylation analysis may have been more appropriate tissue choices. Alternatively, undertaking genome-wide methylation analysis in liver tissue from these mice may more appropriate define regions in the genome that escape epigenetic reprogramming in *MommeD8* mice.

While the expression of Rlf in homozygous *MommeD8* mice is reduced, the activity of Rlf may not be altered in a manner in which large changes in phenotype would be observed. This may result in only subtle alterations in phenotype occurring across multiple generations of *MommeD8* mice. Additionally, the reduced viability of *Rlf* $^{MommeD8/MommeD8}$ mice may suggest these mice are compensating for the loss of *Rlf* in some way or that there is incomplete penetrance in these mice. This may be influencing the results observed in that the *Rlf* $^{MommeD8/MommeD8}$ mice that survive and are used for breeding the next generation are passing on a less severe phenotype than that which may be present in non-viable *Rlf* $^{MommeD8/MommeD8}$ mice. In order to determine whether Rlf has an impact on transgenerational epigenetic inheritance in these mice a greater number of generations would need to be investigated for the accumulation of these subtle effects to be identified with greater specificity.

The results presented here in the hypomorphic mouse lines further support the overall hypothesis of this PhD in that Rlf plays a key role in development and disease progression.

5.4 FUTURE DIRECTIONS

In this chapter I have found a potential role for *Rlf* in ovarian and follicular development. Due to time and financial constraints no studies of how reduced *Rlf* impacts the reproductive system were able to be undertaken. Female and male *Rlf* $^{MommeD8/MommeD34}$ mice have not previously been used for breeding, leaving their fertility status unknown. Crossing *Rlf* $^{MommeD8/MommeD34}$ dams and wild-type sires, and undertaking the reciprocal cross, may reveal whether the differences in ovarian morphology observed impact upon litter size. Investigating follicle development, growth, numbers and serum hormones in females of different ages would more clearly define a time-point or ovarian follicle growth pathways that require *Rlf* for correct development. Alternatively, the production of a conditional Rlf knockout in the ovary may better allow for studying the effect of loss of Rlf rather than reduced expression on ovarian development.

No difference in cardiac structure and function was observed in adult *Rlf* $^{MommeD8/MommeD34}$, however not all *Rlf* $^{MommeD8/MommeD34}$ mice survive to adulthood. This may reflect incomplete phenotypic penetrance amongst compound heterozygous offspring. Western blotting performed in this study was undertaken in adult *Rlf* $^{MommeD8/MommeD34}$ mice, repeating Western blotting in fetal tissue may reveal differences in Rlf protein expression that could point towards differences in Rlf expression amongst compound heterozygous offspring. Similarly, investigating the cardiac function and structure of neonatal or late gestational mice using echocardiography may be able to define which mutants are likely to survive and which would require therapeutic intervention.

Investigation of multiple generations of *MommeD8* homozygous offspring found no significant change in methylation between F1 and F8 mice at both the GFP transgene and *Smad3/Smad6* region in adult liver. As epigenetic reprogramming occurs following fertilisation of the zygote and during the establishment of the primordial germ cells (Hajkova et al., 2002), an earlier time-point may be more beneficial for investigating loci that escape this reprogramming. More suitable loci for investigation may be non-imprinted genes that have been found to inherit methylation from the parental gametes and are thought to escape reprogramming (Borgel et al., 2010). Moving to this early time-point would provide the ability to study mice that are null for *Rlf*. Utilising this mouse line would provide greater detail into the effects of loss of *Rlf* on transgenerational epigenetic inheritance than the hypomorphic *MommeD8* mouse line.

Chapter 6: Overall Discussion and Future Directions

This is the first study to investigate the molecular and phenotypic consequences of loss of Rlf in the mouse. The results presented here have directly addressed the main hypothesis of this PhD. That *Rlf* is a novel epigenetic modifier of critical importance in development and disease.

The research undertaken here has investigated three key aims which will be addressed below.

Aim 1: Loss of Rlf alters the transcriptome and methylome in E14.5 fetal mouse liver

Both the methylome and transcriptome were found to be altered in *Rlf* null fetal liver. RNA-seq identified a number of genes that were both up- and down-regulated in E14.5 liver. More genes were found to be down-regulated, supporting a role for Rlf in transcriptional activation, and these genes displayed higher fold changes in expression than those that were up-regulated. ChIPseq, using two independent Rlf antibodies, also identified putative Rlf binding sites located at or near TSS. Further supporting a role for *Rlf* in transcriptional activation.

Methylation in Rlf null fetal liver was found to be increased at a number of distinct loci across the genome. Many of these differentially methylated regions overlapped with putative regulatory elements, although these did not overlap with putative Rlf binding sites. This overlap is reminiscent of tissue specific DMRs that retain an epigenetic memory from early in development. Our results suggest that Rlf

is also involved in the maintenance of hypomethylation at such regions. It is thought tissue specific differential methylation of CpG island shores may be transcription factor dependent, and this is supported by our findings that loss of Rlf results in increased methylation of CpG island shores.

Here I have proposed a model that Rlf binding to TSS may stabilise the binding of an independent factor to the Rlf-DMR, facilitating the demethylation and expression of these genes. Further study, utilising methods such as chromosome conformation capture, will help to uncover the precise mechanism through which Rlf influences gene expression.

Aim 2: Rlf is of critical importance in mouse development

Here I have shown that Rlf is expressed widely during development, and mice with reduced or no expression of Rlf are not present at adulthood in the expected numbers. Homozygous *MommeD28* and *MommeD34* mid-gestation (E14.5) embryos were found to display defects in cardiac development, reminiscent of left ventricular non-compaction (LVNC) observed in humans. RNA-seq performed in E13.5 hearts, prior to the observation of the heart defect, revealed small but significant changes in the expression of genes that are implicated in pathways that are critical for cardiac development. For example, a number of dysregulated genes form a part of the Notch1 signalling pathway which is essential for compaction of the heart wall. This dysregulated signalling, as a result of down regulation of Notch1 ligands, may lead to the cardiac defect observed. Future use of these first *Rlf* mouse mutants may allow for a greater understanding of LVNC and how it arises.

As the *Momme* mice used in this study display global loss of Rlf expression, the development of conditional Rlf knockouts in cardiac specific cells, e.g. endocardial cells, may confirm whether the cardiac defect observed in these mice is specific to loss of Rlf in one cell type or multiple cell types within the heart. Additionally, no studies have previously linked Rlf with LVNC in humans; undertaking sequencing of clinical samples in the future may enhance current screening, diagnosis and management strategies for these patients.

The placenta also plays an important role in regulating cardiovascular growth and fetal development. Here I have identified subtle defects in placental fetal vasculature in *Rlf* mutants at E14.5. This difference in morphology from mid-gestation may be responsible for the reduction in weight observed in late gestation *Rlf* null embryos. Aggregating wild-type embryos with *Rlf* mutant placentas, or the reciprocal aggregation, may be able to more clearly define the relationship between the cardiac defect, weight reduction and placental defect observed.

Aim 3: Reduced Rlf expression alters phenotype in the adult mouse

I have shown hypomorphic $Rlf^{MommeD8/MommeD8}$ and $Rlf^{MommeD8/MommeD34}$ mice, with reduced *Rlf* expression, have a greater impact on endogenous gene expression than $Rlf^{MommeD28/+}$ and $Rlf^{MommeD34/+}$ mice but not as great an effect as $Rlf^{MommeD28/MommeD28}$ and $Rlf^{MommeD34/MommeD34}$ mice. This suggests the different *Rlf* *Momme* mutants have varying impacts on *Rlf* activity within the mouse. The availability of these different *Rlf* mouse mutants provide a unique tool with which to study the effect of reduced *Rlf* expression on phenotype that may reflect occurrences in the human population.

Hypomorphic *Rlf* mutants were used to study the effect of loss of Rlf on phenotype across generations. No significant accumulation of phenotype was observed in these mice, suggesting the epigenetic state is reset between each generation. The methods used in this study however, would not have been sufficient for detecting subtle changes in phenotype. Additionally, only two loci were investigated that may not have been representative of the whole genome or loci that escape reprogramming. Future studies investigating an earlier developmental time point, for example following the fertilisation of the zygote or imprint erasure in primordial germ cells, may determine whether *Rlf* affects the clearing of DNA methylation between generations.

Differences in ovarian morphology were also observed in a preliminary cohort of hypomorphic *Rlf* female mice. The absence of oocytes and follicles may arise from reduced *Rlf* expression early in development, when primordial follicles are being established. Future investigations of the developing follicles during embryogenesis may reveal a critical role for *Rlf* in follicle development, which may in turn influence fertility and/or tumorigenic transformation.

The use of these first known mouse mutants for *Rlf* in future investigations will be valuable in advancing our knowledge of cardiac development, epigenetic reprogramming and ovarian development.

Chapter 7: Appendices

Appendix I: PCR Primer Sequences

PCR primers for genotyping

Primer Name	Forward Sequence (5' to 3')	Reverse Sequence (5' to 3')
MommeD8	GCAGCCAGAAAGTAGCAAATGAGCG	TGTTTGGGAGGGGCACACTGC
MommeD28	TCTGCCAGTCTCTGAAGAAAAGGCA	CTGTGGCTTTCATTCTGGAAGGGAA
MommeD34	GTTGCCCATTTTAGGGGAG	TGGATTGCACCCTGGACTAC

Primers for qRT-PCR

Primer Name	Forward Sequence (5' to 3')	Reverse Sequence (5' to 3')
Hpd	AGGTAGTCAGCCACGTCATC	CAATGTGGTCGCAGTCTTCC
Timd2	ATGGTGGTCCCTATTGCTGT	GGCCTCTGGTTTGTAGGTCT
Apoc1	CATCGCTCTTCCTGTCCTGA	ATATGTTCAATGGCTGCCCG
Prss50	GGTTCATTCCAGCAACCTCC	GAAGCGATAAGGATGCCAGC
Atp2b2	CCTCAAAACCTCGCCTGTTG	GTGGGTGGTAGAAGGACAGT
Crispld2	GGTACGACGAGGTGAAGGAT	TACACGGCATTCTCCCAAGT
Psma8	ACTGTGAGGAAAATCTGCGC	TACCAAAAGGCCTTCGTCCA
Upb1	AGGAATCTCGATCTGCCCAG	ATTGACTCCACACATTGCGG
Mgam	TTGTTCTGCTGCTTGTCCTG	ACTGGGCAATTGGGAGAGTT
Myo5c	GGCTGAAATCGCAAAGGACT	CTCATGGAGGTAGCTGAGGG
Vldlr	TCGGGCTTTGTTTACTGGTC	AGTAGAGGCGGCTTTTGACA
Aldh1a7	GCAGGGAAAAGCAATCTGAA	TCTGACCCTGGTGGAAGAAC

Primers for Bisulphite Sequencing

Primer Name	Sequence (5' to 3')	Tm (°C)	Product Size
Smad3/6 External Forward	AAGTGGAATTTTTTAGTGGTAGATG	53	493
Smad3/6 External Reverse	AACTACTTTAATAAAAAATAACATAACC	53	493
Smad3/6 Internal Forward	TTGGTATGTGTTGTTTTTAGTTTTG	51	493
Smad3/6 Internal Reverse	ACAATTTAACTATTCATTATATCTCTAACA	51	493
Hpd External Forward	TTTAGAATTTAGAGGTAGTAGGGTAAAT	60	300
Hpd External Reverse	ACAAAATATACCCTAACACTCAACTCAT	60	300
Hpd Internal Forward	GGATTTTAAATGATAGTTGGGAGATTTT	60	300
Hpd	CCTAACCAAAACTCAAATCACATCTA	60	300

Primer Name	Sequence (5' to 3')	Tm (°C)	Product Size
Internal Reverse			
Prss50 External Forward	GGTATAGGGATTTAGAGAAAGGG	60	326
Prss50 External Reverse	TTAGGAGTTTGAGGAAAGATATAAGGTA	60	326
Prss50 Internal Forward	ACCAAATAACCTATAAAAAATCTCCA	60	326
Prss50 Internal Reverse	ACCTAAAAAACAAATACCTCTAAAACC	60	326
Basp1 External Forward	GAGGTTAAAAAGATTGAGGTTTT	47	309
Basp1 External Reverse	TCCAATTTAAAACAAATTAATATAAAATAA	47	309
Basp1 Internal Forward	GTTAAGGTAGAGGAGAAGGAGGT	47	309
Basp1 Internal Reverse	AATAAAACACATCCTCTTTATTTTT	47	309
GFP enhancer External Forward	AAAATAAAATTTTTGGATTGTTATTATTATAA	60	162
GFP enhancer External/Internal Reverse	AATCTCTACTCACTACAAACTCCATCTC	60	162
GFP enhancer Internal Forward	ATATTTGTAATTTTAGTATTTTGGGAGGTT	60	162
Enhancer mm87 External Forward	GGTAGGAGATAATGGGTTTAAAGGT	56	350
Enhancer mm87 External Reverse	TCAACTCCCAAAAAATAAACTAAAC	56	350
Enhancer mm87 Internal Forward	TGGGTTTAAAGGTGTTGAATTAGTT	56	350
Enhancer mm87 Internal Reverse	TAAATAAAATAAAACCCACTCCTCCC	56	350
SP6	TATTTAGGTGACACTATAG	55	Variable
T7	TAATACGACTCACTATAGGG	55	Variable

Appendix II: Differential expression analysis comparing RNA-seq datasets from *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ E14.5 liver.

Please view the accompanying Excel spreadsheet of the DESeq Differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ mice to *Rlf*$^{MommeD28/MommeD28}$ mice. The column headings represent:

Column Header	Information
Ensembl gene ID	Ensembl identification number
Gene Name	Short gene name (symbol)
baseMean	The mean of the counts divided by the size factors for the counts for both conditions (wild-type and homozygous)
baseMeanA	The base mean i.e., mean of the counts divided by the size factors for the counts for *Rlf*$^{+/+}$
baseMeanB	The base mean for i.e., mean of the counts divided by the size factors for the counts for *Rlf*$^{MommeD28/MommeD28}$
FoldChange	Ratio meanB/meanA
Log2FoldChange	Log2 of the fold change
Pval	p value for rejecting the null hypothesis 'meanA==meanB'
padj	adjusted p values (adjusted with 'p.adjust)

Appendix III: Differential expression analysis comparing RNA-seq datasets from *Rlf*$^{+/+}$ and *Rlf*$^{MommeD28/+}$ E14.5 liver.

Please view the accompanying Excel spreadsheet of the DESeq Differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ mice to *Rlf*$^{MommeD28/+}$ mice. The column headings represent:

Column Header	Information
Ensembl gene ID	Ensembl identification number
Gene Name	Short gene name (symbol)
baseMean	The mean of the counts divided by the size factors for the counts for both conditions (wild-type and heterozygous)
baseMeanA	The base mean i.e., mean of the counts divided by the size factors for the counts for *Rlf*$^{MommeD28/+}$
baseMeanB	The base mean for i.e., mean of the counts divided by the size factors for the counts for *Rlf*$^{+/+}$
FoldChange	Ratio meanB/meanA
Log2FoldChange	Log2 of the fold change
Pval	p value for rejecting the null hypothesis 'meanA==meanB'
padj	adjusted p values (adjusted with 'p.adjust)

Appendix IV: Rlf-DMRs identified using genome-wide bisulphite sequencing of DNA from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ E14.5 liver.

Please view the accompanying Excel spreadsheet of the GWBS differential methylated analysis comparing *Rlf* $^{+/+}$ to *Rlf* $^{MommeD28/MommeD28}$ fetal liver. The column headings represent:

Column Header	Information
Chr	Chromosome number
Start	Start coordinate of Rlf-DMR
End	End coordinate of Rlf-DMR
CpGs	Number of CpGs within Rlf-DMR
mCG difference	Difference in methylation between genotypes
Rlf $^{+/+}$ mean mCG	Mean methylation for wild-type replicates
Rlf $^{MommeD28/MommeD28}$ mean mCG	Mean methylation for the homozygous replicates
Direction	Direction in which methylation is occurring in *Rlf* $^{MommeD28/MommeD28}$ fetal liver

Appendix V: Rlf peaks detected in chromatin extracted from *Rlf* $^{+/+}$ E14.5 liver following ChIPseq with Abcam Rlf antibody.

Please view the accompanying Excel spreadsheet ChIPseq analysis of *Rlf* $^{+/+}$ fetal liver following ChIP with the Abcam Rlf antibody. The column headings represent:

Column Header	Information
Active Region	Unique identifier assigned to each Active Region
Chromosome	Chromosome number
Start	Nucleotide position on chromosome where Active Region starts (if on + strand) or ends (if on – strand)
End	Nucleotide position on chromosome where gene ends (if on + strand) or starts (if on – strand)
Length	Length of Active Region in base pairs
IntervalCount	Number of Intervals that make up the Active Region
Wild_1_AvgValue	Value Average of fragment densities of all the bins within the Active Region
Wild_2_AvgValue	Value Average of fragment densities of all the bins within the Active Region
Wild_1_Peak Value	Peak Value Peak fragment density of the Active Region (Value at "Summit")
Wild_2_Peak Value	Peak Value Peak fragment density of the Active Region (Value at "Summit")
CGIslandCount	Number of UCSC-annotated CpG Islands within a set distance of each Active Region
Promoter Count	Number of Promoters (-7500 to +2500 bp relative to gene starts) overlapping with each Active Region (only for tables with <65,530 rows)
GeneCount	Number of NCBI annotated genes within a set distance (GeneMargin) of each Active Region
Gene List	Names (symbols) of genes found within a set distance (GeneMargin) of Active Region
Dist to Start	Distance of midpoint of Active Region to gene start
Position	Location of midpoint of Active Region relative to gene
UCSC Link	Link to display Active Region +/- 0.5 kb flanking sequence in browser
Wild_1Present	Present Presence of an Interval from [Sample x] in an Active Region is denoted by "1" and absence is denoted by "0". These columns are useful for sorting data to determine patterns between samples
Wild_2 Present	Present Presence of an Interval from [Sample x] in an Active Region is denoted by "1" and absence is denoted by "0". These columns are useful for sorting data to determine patterns between samples

Appendix VI: Rlf peaks detected in chromatin extracted from *Rlf* $^{+/+}$ E14.5 liver following ChIPseq with a custom Rlf antibody.

Please view the accompanying Excel spreadsheet ChIPseq analysis of *Rlf* $^{+/+}$ fetal liver following ChIP with the Abcam Rlf antibody. The column headings represent:

Column Header	Information
Active Region	Unique identifier assigned to each Active Region
Chromosome	Chromosome number
Start	Nucleotide position on chromosome where Active Region starts (if on + strand) or ends (if on – strand)
End	Nucleotide position on chromosome where gene ends (if on + strand) or starts (if on – strand)
Length	Length of Active Region in base pairs
IntervalCount	Number of Intervals that make up the Active Region
Wild_5_AvgValue	Value Average of fragment densities of all the bins within the Active Region
Wild_6_AvgValue	Value Average of fragment densities of all the bins within the Active Region
Wild_5_Peak Value	Peak Value Peak fragment density of the Active Region (Value at "Summit")
Wild_6_Peak Value	Peak Value Peak fragment density of the Active Region (Value at "Summit")
CGIslandCount	Number of UCSC-annotated CpG Islands within a set distance of each Active Region
Promoter Count	Number of Promoters (-7500 to +2500 bp relative to gene starts) overlapping with each Active Region (only for tables with <65,530 rows)
GeneCount	Number of NCBI annotated genes within a set distance (GeneMargin) of each Active Region
Gene List	Names (symbols) of genes found within a set distance (GeneMargin) of Active Region
Dist to Start	Distance of midpoint of Active Region to gene start
Position	Location of midpoint of Active Region relative to gene
UCSC Link	Link to display Active Region +/- 0.5 kb flanking sequence in browser
Wild_5Present	Present Presence of an Interval from [Sample x] in an Active Region is denoted by "1" and absence is denoted by "0". These columns are useful for sorting data to determine patterns between samples
Wild_6 Present	Present Presence of an Interval from [Sample x] in an Active Region is denoted by "1" and absence is denoted by "0". These columns are useful for sorting data to determine patterns between samples

Appendix VII: Histology sections of cohorts used for defining cardiac defect in *Rlf* mutant mice.

Please view the accompanying PDF file of the E14.5, E9.5 and E11.5 *Rlf* $^{+/+}$ and *Rlf*$^{MommeD28/MommeD28}$ cohorts used for histological analysis of the heart.

Appendix VIII: CuffDiff Differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/+}$ E13.5 hearts.

Please view the accompanying Excel spreadsheet of the CuffDiff differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal hearts. The column headings represent:

Column Header	Information
Gene	The gene_name(s) or gene_id(s) being tested
Ensembl gene ID	Ensembl identification number
Locus	Genomic coordinates for easy browsing to the genes or transcripts being tested.
Sample 1	Label (or number if no labels provided) of the first sample being tested
Sample 2	Label (or number if no labels provided) of the second sample being tested
Test status	Can be one of OK (test successful), NOTEST (not enough alignments for testing), LOWDATA (too complex or shallowly sequenced), HIDATA (too many fragments in locus), or FAIL, when an ill-conditioned covariance matrix or other numerical exception prevents testing.
Mean Wild-type	FPKM of the gene in sample wild-types replicates
Mean Homozygous	FPKM of the gene in sample homozygous replicates
FoldChange	Ratio MeanHomozygous/meanWild-type
log2(FPKMy/FPKMx)	The (base 2) log of the fold change y/x
Test stat	The value of the test statistic used to compute significance of the observed change in FPKM
p	The uncorrected p-value of the test statistic
q	The FDR-adjusted p-value of the test statistic
Significant	Can be either “yes” or “no”, depending on whether p is greater than the FDR after Benjamini-Hochberg correction for multiple-testing

Appendix IX: EdgeR Differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/+}$ E13.5 hearts.

Please view the accompanying Excel spreadsheet of the EdgeR differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal hearts. The column headings represent:

Column Header	Information
Gene name	The gene_name(s) or gene_id(s) being tested
Ensembl gene ID	Ensembl identification number
Sample (x)	Counts per million (CPM) mapped reads scaled by the number of fragments sequenced for each replicate
BaseMean Homo	Mean of the counts divided by the size factors for the counts for *Rlf* $^{MommeD28/+}$
BaseMean Wild	Mean of the counts divided by the size factors for the counts for *Rlf* $^{+/+}$
FoldChange	Ratio meanB/meanA
Log2FoldChange	Log2 of the fold change
p val	p value for rejecting the null hypothesis 'meanHomo==meanWild'
Adj p val	adjusted p values

Appendix X: DESeq Differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/+}$ E13.5 hearts.

Please view the accompanying Excel spreadsheet of the DESeq differential expression analysis comparing RNA-seq datasets from *Rlf* $^{+/+}$ and *Rlf* $^{MommeD28/MommeD28}$ fetal hearts. The column headings represent:

Column Header	Information
Gene name	The gene_name(s) or gene_id(s) being tested
Ensembl gene ID	Ensembl identification number
baseMean	The mean of the counts divided by the size factors for the counts for both conditions (wild-type and heterozygous)
FoldChange	Ratio meanB/meanA
Log2FoldChange	Log2 of the fold change
lfcSE	Standard error: *Rlf* $^{MommeD28/MommeD28}$:*Rlf* $^{+/+}$
Stat	Wald statistic: *Rlf* $^{MommeD28/MommeD28}$:*Rlf* $^{+/+}$
pval	p value for rejecting the null hypothesis 'meanA==meanB'
padj	adjusted p values (adjusted with 'p.adjust)
NReads	

Appendix XI: Histological sections of aged female *Rlf*$^{+/+}$ and *Rlf*$^{MommeD8/MommeD34}$ ovaries.

Please view the accompanying PDF file of the histological ovarian sections from aged female cohort.

Appendix XII: Permission for Figure 1.2

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 13, 2016

This is a License Agreement between Lauren M Bourke ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Lauren M Bourke
Customer address	300 Herston Rd Brisbane, 4051
License number	3787340948860
License date	Jan 13, 2016
Licensed content publisher	Elsevier
Licensed content publication	Cell
Licensed content title	Transgenerational Epigenetic Inheritance: Myths and Mechanisms
Licensed content author	Edith Heard,Robert A. Martienssen
Licensed content date	27 March 2014
Licensed content volume number	157
Licensed content issue number	1
Number of pages	15
Start Page	95
End Page	109
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

Original figure numbers	Figure 2
Title of your thesis/dissertation	Investigating the role of a novel epigenetic modifier, Rearranged L-Myc Fusion, Rlf
Expected completion date	Feb 2016
Estimated size (number of pages)	240
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 AUD
VAT/Local Sales Tax	0.00 AUD / 0.00 GBP
Total	0.00 AUD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and

accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the

Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.
Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:
- immediately
via their non-commercial person homepage or blog
by updating a preprint in arXiv or RePEc with the accepted manuscript
via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
directly by providing copies to their students or to research collaborators for their personal use
for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
via non-commercial hosting platforms such as their institutional repository
via commercial sites with which Elsevier has an agreement
In all cases accepted manuscripts should:
- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors the following clauses are applicable in addition to the above:** Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in

Appendix XIII: Permission Figure 1.3

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Jan 13, 2016

This is a License Agreement between Lauren M Bourke ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3787260468331
License date	Jan 13, 2016
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Genetics
Licensed content title	Epigenetic inheritance at the agouti locus in the mouse
Licensed content author	Hugh D. Morgan, Heidi G.E. Sutherland, David I.K. Martin, Emma Whitelaw
Licensed content date	Nov 1, 1999
Volume number	23
Issue number	3
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Figures	Figure 1
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Investigating the role of a novel epigenetic modifier, Rearranged L-Myc Fusion, Rlf
Expected completion date	Feb 2016
Estimated size (number of pages)	240
Total	0.00 AUD
Terms and Conditions	

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the

material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.
3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free)
4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.
5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer*, the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)
6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)
7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer*, the following credit

line applies.
Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix XIV: Permission for Figure 3.4

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 13, 2016

This is a License Agreement between Lauren M Bourke ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Lauren M Bourke
Customer address	300 Herston Rd Brisbane, 4051
License number	3787360904673
License date	Jan 13, 2016
Licensed content publisher	Elsevier
Licensed content publication	Molecular Genetics and Metabolism
Licensed content title	Gene expression profiles of homogentisate-treated Fah−/− Hpd−/−mice using DNA microarrays
Licensed content author	Yasuhiko Tanaka,Kimitoshi Nakamura,Shirou Matsumoto,Yoshiko Kimoto,Akito Tanoue,Gozoh Tsujimoto,Fumio Endo
Licensed content date	November 2006
Licensed content volume number	89
Licensed content issue number	3
Number of pages	7
Start Page	203
End Page	209
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Investigating the role of a novel epigenetic modifier, Rearranged L-Myc Fusion, Rlf

Expected completion date	Feb 2016
Estimated size (number of pages)	240
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 AUD
VAT/Local Sales Tax	0.00 AUD / 0.00 GBP
Total	0.00 AUD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:
- immediately
via their non-commercial person homepage or blog
by updating a preprint in arXiv or RePEc with the accepted manuscript
via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
directly by providing copies to their students or to research collaborators for their personal use
for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
via non-commercial hosting platforms such as their institutional repository
via commercial sites with which Elsevier has an agreement
In all cases accepted manuscripts should:
- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.
Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors the following clauses are applicable in addition to the above:**
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you

scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.
Additional Terms & Conditions applicable to each Creative Commons user license:
CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.
CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.
Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.
Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Appendix XV: Permission for Figure 4.14

ELSEVIER LICENSE
TERMS AND CONDITIONS
Jan 31, 2016

This is a License Agreement between Lauren M Bourke ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Lauren M Bourke
Customer address	300 Herston Rd Brisbane, 4051
License number	3799550985928
License date	Jan 31, 2016
Licensed content publisher	Elsevier
Licensed content publication	Developmental Biology
Licensed content title	Interactions between Trophoblast Cells and the Maternal and Fetal Circulation in the Mouse Placenta
Licensed content author	S.Lee Adamson,Yong Lu,Kathie J. Whiteley,Doug Holmyard,Myriam Hemberger,Christine Pfarrer,James C. Cross
Licensed content date	15 October 2002
Licensed content volume number	250
Licensed content issue number	2
Number of pages	16
Start Page	358
End Page	373
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 10
Title of your thesis/dissertation	Investigating the role of a novel epigenetic modifier, Rearranged L-Myc Fusion, Rlf
Expected completion date	Feb 2016
Estimated size (number of pages)	240
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 AUD

VAT/Local Sales Tax	0.00 AUD / 0.00 GBP
Total	0.00 AUD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized

pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.
LIMITED LICENSE
The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal

publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

– immediately

via their non-commercial person homepage or blog

by updating a preprint in arXiv or RePEc with the accepted manuscript

via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group

directly by providing copies to their students or to research collaborators for their personal use

for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement

– after the embargo period

via non-commercial hosting platforms such as their institutional repository

via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

– link to the formal publication via its DOI

– bear a CC-BY-NC-ND license - this is easy to do

– if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors the following clauses are applicable in addition to the above:**

Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be

Chapter 8: References

Adamson, S. L., Lu, Y., Whiteley, K. J., Holmyard, D., Hemberger, M., Pfarrer, C., & Cross, J. C. (2002). Interactions between trophoblast cells and the maternal and fetal circulation in the mouse placenta. *Developmental Biology, 250*(2), 358-373. doi:10.1006/dbio.2002.0773

Allis, C. D., Jenuwein, T., & Reinberg, D. (2007). *Epigenetics*. Cold Spring Harbour, N.Y: Cold Spring Laboratory Press.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology, 11*(10), R106. doi:10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., & Huber, W. (2015). Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-169. doi:10.1093/bioinformatics/btu638

Andrews, S. (2010). Fastqc: A quality control tool for high throughput sequencing data. from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Angulo, M., Butler, M., & Cataletto, M. (2015). Prader-willi syndrome: A review of clinical, genetic, and endocrine findings. *Journal of endocrinological investigation*, 1-15. doi:10.1007/s40618-015-0312-9

Ashe, A., Morgan, D. K., Whitelaw, N. C., Bruxner, T. J., Vickaryous, N. K., Cox, L. L., . . . Whitelaw, E. (2008). A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biology, 9*(12), R182. doi:10.1186/gb-2008-9-12-r182

Auer, P. L., & Doerge, R. W. (2010). Statistical design and analysis of rna sequencing data. *Genetics, 185*(2), 405-416. doi:10.1534/genetics.110.114983

Australian Institute of Health and Welfare. (2011). *Cardiovascular disease: Australian facts 2011. Cvd series no.35. Cat. No. Cvd 53.* Canberra: AIHW.

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., . . . Zhang, J. (2013). Practical guidelines for the comprehensive analysis of chip-seq data. *PLoS Computational Biology, 9*(11), e1003326. doi:10.1371/journal.pcbi.1003326

Banerjee, I., Fuseler, J. W., Price, R. L., Borg, T. K., & Baudino, T. A. (2007). Determination of cell types and numbers during cardiac development in the neonatal and adult rat and mouse. *American Journal of Physiology - Heart and Circulatory Physiology, 293*(3), H1883-H1891. doi:10.1152/ajpheart.00514.2007

Barak, Y., Nelson, M. C., Ong, E. S., Jones, Y. Z., Ruiz-Lozano, P., Chien, K. R., . . . Evans, R. M. (1999). Pparγ is required for placental, cardiac, and adipose tissue development. *Molecular cell, 4*(4), 585-595. doi:10.1016/S1097-2765(00)80209-9

Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K., & Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the tme locus. *Nature, 349*(6304), 84-87. doi:10.1038/349084a0

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837. doi:10.1016/j.cell.2007.05.009

Bartolomei, M. S., & Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology, 3*(7). doi:10.1101/cshperspect.a002592

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300

Bestor, T. H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a zn binding regulatory domain. *The EMBO Journal, 11*(7), 2611

Bevilacqua, P. C., & Blose, J. M. (2008). Structures, kinetics, thermodynamics, and biological functions of rna hairpins. *Annual Review of Physical Chemistry, 59*(1), 79-103. doi:10.1146/annurev.physchem.59.032607.093743

Birnbaum, R. Y., Clowney, E. J., Agamy, O., Kim, M. J., Zhao, J., Yamanaka, T., . . . Ahituv, N. (2012). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res, 22*(6), 1059-1068. doi:10.1101/gr.133546.111

Blankenberg, D., Kuster, G. V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., . . . Taylor, J. (2001). Galaxy: A web-based genome analysis tool for experimentalists *Current Protocols in Molecular Biology*: John Wiley & Sons, Inc.

Blewitt, M. E., Gendrel, A. V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., . . . Whitelaw, E. (2008). Smchd1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in x inactivation. *Nature Genetics, 40*(5), 663-669. doi:10.1038/ng.142

Blewitt, M. E., Vickaryous, N. K., Hemley, S. J., Ashe, A., Bruxner, T. J., Preis, J. I., . . . Whitelaw, E. (2005). An n-ethyl-n-nitrosourea screen for genes involved in variegation in the mouse. *Proceedings of National Academy of Sciences U S A, 102*(21), 7629-7634. doi:10.1073/pnas.0409375102

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., & Lengauer, T. (2005). Biq analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics, 21*(21), 4067-4068. doi:10.1093/bioinformatics/bti652

Borden, M., Holm, J., Leslie, J., Sweetman, L., Nyhan, W. L., Fleisher, L., . . . Scott, C. R. (1992). Hawkinsinuria in two families. *American Journal of Medical Genetics, 44*(1), 52-56. doi:10.1002/ajmg.1320440113

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., . . . Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet, 42*(12), 1093-1100. doi:http://www.nature.com/ng/journal/v42/n12/abs/ng.708.html#supplementary-information

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). Uhrf1 plays a role in maintaining DNA methylation in mammalian cells. *Science, 317*(5845), 1760-1764. doi:10.1126/science.1147939

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., . . . Rastan, S. (1992). The product of the mouse xist gene is a 15 kb inactive x-specific transcript containing no conserved orf and located in the nucleus. *Cell, 71*(3), 515-526

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human xist gene: Analysis of a 17 kb inactive x-specific rna that contains conserved repeats and is highly localized within the nucleus. *Cell, 71*(3), 527-542. doi:10.1016/0092-8674(92)90520-M

Bultman, S. J., Michaud, E. J., & Woychik, R. P. (1992). Molecular characterization of the mouse agouti locus. *Cell, 71*(7), 1195-1204. doi:10.1016/S0092-8674(05)80067-4

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., . . . Zhang, Y. (2002). Role of histone h3 lysine 27 methylation in polycomb-group silencing. *Science, 298*(5595), 1039-1043. doi:10.1126/science.1076997

Castrillon, D. H., Miao, L., Kollipara, R., Horner, J. W., & DePinho, R. A. (2003). Suppression of ovarian follicle activation in mice by the transcription factor foxo3a. *Science, 301*(5630), 215-218. doi:10.1126/science.1086336

Cerone, R., Holme, E., Schiaffino, M. C., Caruso, U., Maritano, L., & Romano, C. (1997). Tyrosinemia type iii: Diagnosis and ten-year follow-up. *Acta Paediatrica, 86*(9), 1013-1015. doi:10.1111/j.1651-2227.1997.tb15192.x

Chahoud, I., & Paumgartten, F. J. R. (2009). Influence of litter size on the postnatal growth of rat pups: Is there a rationale for litter-size standardization in toxicity studies? *Environmental Research, 109*(8), 1021-1027. doi:10.1016/j.envres.2009.07.015

Chai, J., Charboneau, A. L., Betz, B. L., & Weissman, B. E. (2005). Loss of the hsnf5 gene concomitantly inactivates p21cip/waf1 and p16ink4a activity associated with replicative senescence in a204 rhabdoid tumor cells. *Cancer Research, 65*(22), 10192-10198. doi:10.1158/0008-5472.can-05-1896

Chakravarthy, S., & Luger, K. (2006). The histone variant macro-h2a preferentially forms “hybrid nucleosomes”. *Journal of Biological Chemistry, 281*(35), 25522-25531. doi:10.1074/jbc.M602258200

Chang, C. P., & Bruneau, B. G. (2012). Epigenetics and cardiovascular development. *Annual review of physiology, 74*, 41-68. doi:10.1146/annurev-physiol-020911-153242

Changolkar, L. N., & Pehrson, J. R. (2002). Reconstitution of nucleosomes with histone macroh2a1. 2. *Biochemistry, 41*(1), 179-184. doi:10.1021/bi0157417

Chen, C.-C., Wang, K.-Y., & Shen, C.-K. J. (2012). The mammalian de novo DNA methyltransferases dnmt3a and dnmt3b are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *Journal of Biological Chemistry, 287*(40), 33116-33121. doi:10.1074/jbc.C112.406975

Chen, H., Zhang, W., Sun, X., Yoshimoto, M., Chen, Z., Zhu, W., . . . Shou, W. (2013). Fkbp1a controls ventricular myocardium trabeculation and compaction by regulating endocardial notch1 activity. *Development, 140*(9), 1946-1957. doi:10.1242/dev.089920

Cheung, P., & Lau, P. (2005). Epigenetic regulation by histone methylation and histone variants. *Molecular Endocrinology, 19*(3), 563-573. doi:10.1210/me.2004-0496#sthash.l9faYQvE.dpuf

Chien, P. F. W. (1991). Investigations of protein metabolism in human pregnancy: The term foetus and placenta studied using stable isotope labelled amino-acids. *Clinical Nutrition, 10, Supplement*, 70-76. doi:10.1016/0261-5614(91)90119-W

Chong, S., Vickaryous, N., Ashe, A., Zamudio, N., Youngson, N., Hemley, S., . . . Whitelaw, E. (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nature Genetics, 39*(5), 614-622. doi:10.1038/ng2031

Chowdhury, S., Erickson, S. W., MacLeod, S. L., Cleves, M. A., Hu, P., Karim, M. A., & Hobbs, C. A. (2011). Maternal genome-wide DNA methylation patterns and congenital heart defects. *PLoS One, 6*(1), e16506. doi:10.1371/journal.pone.0016506

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., . . . Chang, H. Y. (2015). Systematic discovery of xist rna binding proteins. *Cell, 161*(2), 404-416. doi:10.1016/j.cell.2015.03.025

Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annual review of biochemistry, 78*, 273-304. doi:10.1146/annurev.biochem.77.062706.153223

Clark, S. J., Harrison, J., Paul, C. L., & Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res, 22*(15), 2990-2997

Conway, S. J., Kruzynska-Frejtag, A., Kneer, P. L., Machnicki, M., & Koushik, S. V. (2003). What cardiovascular defect does my prenatal mouse mutant have, and why? *genesis, 35*(1), 1-21. doi:10.1002/gene.10152

Corstius, H. B., Zimanyi, M. A., Maka, N., Herath, T., Thomas, W., van der Laarse, A., . . . Black, M. J. (2005). Effect of intrauterine growth restriction on the number of cardiomyocytes in rat hearts. *Pediatr Res, 57*(6), 796-800

Cropley, J. E., Suter, C. M., Beckman, K. B., & Martin, D. I. K. (2006). Germ-line epigenetic modification of the murine avy allele by nutritional supplementation. *Proceedings of the National Academy of Sciences, 103*(46), 17308-17312. doi:10.1073/pnas.0607090103

Cropley, J. E., Suter, C. M., Beckman, K. B., & Martin, D. I. K. (2010). Cpg methylation of a silent controlling element in the murine avy allele is

incomplete and unresponsive to methyl donor supplementation. *PLoS One, 5*(2), e9055. doi:10.1371/journal.pone.0009055

D'Amato, G., Luxan, G., del Monte-Nieto, G., Martinez-Poveda, B., Torroja, C., Walter, W., . . . de la Pompa, J. L. (2016). Sequential notch activation regulates ventricular chamber development. *Nat Cell Biol, 18*(1), 7-20. doi:10.1038/ncb3280

http://www.nature.com/ncb/journal/v18/n1/abs/ncb3280.html#supplementary-information

Dai, Z., Dai, X., Xiang, Q., Feng, J., Wang, J., Deng, Y., & He, C. (2009). Genome-wide analysis of interactions between atp-dependent chromatin remodeling and histone modifications. *BioMed Central Genomics, 10*, 304. doi:10.1186/1471-2164-10-304

Danilovich, N., & Sairam, M. R. (2002). Haploinsufficiency of the follicle-stimulating hormone receptor accelerates oocyte loss inducing early reproductive senescence and biological aging in mice. *Biology of reproduction, 67*(2), 361-369. doi:10.1095/biolreprod67.2.361

Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A., . . . Stark, R. (2008). Piwi and pirnas act upstream of an endogenous sirna pathway to suppress tc3 transposon mobility in the caenorhabditis elegans germline. *Molecular cell, 31*(1), 79-90. doi:10.1016/j.molcel.2008.06.003

Daxinger, L., Harten, S., Oey, H., Epp, T., Isbel, L., Huang, E., . . . Whitelaw, E. (2013). An enu mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biology, 14*(9), R96. doi:10.1186/gb-2013-14-9-r96

Dean, W., Santos, F., & Reik, W. (2003). Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. *Seminars in Cell & Developmental Biology, 14*(1), 93-100. doi:10.1016/S1084-9521(02)00141-6

Delmas, V., Stokes, D., & Perry, R. (1993). A mammalian DNA-binding protein that contains a chromodomain and an snf2/swi2-like helicase domain. *Proceedings of the National Academy of Sciences, 90*(6), 2414-2418

Dupont, C., Armant, D. R., & Brenner, C. A. (2009). Epigenetics: Definition, mechanisms and clinical perspective. *Seminars in Reproductive Medicine, 27*(5), 351-357. doi:10.1055/s-0029-1237423

Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M., & Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Molecular cell, 49*(2), 368-378

Eden, S., Hashimshony, T., Keshet, I., Cedar, H., & Thorne, A. (1998). DNA methylation models histone acetylation. *Nature, 394*(6696), 842-842

Elbashir, S. M., Lendeckel, W., & Tuschl, T. (2001). Rna interference is mediated by 21- and 22-nucleotide rnas. *Genes & Development, 15*(2), 188-200. doi:10.1101/gad.862301

Elgin, S. C. R., Luu, Y., Kolasinska-Zwierz, P., Gorchakov, A. A., Egelhofer, T. A., Vielle, A., . . . Kharchenko, P. V. (2011). An assessment of histone-modification antibody quality. *Nature structural & molecular biology, 18*(1), 91-93. doi:10.1038/nsmb.1972

Ellaway, C. J., Holme, E., Standing, S., Preece, M. A., Green, A., Ploechl, E., . . . Leonard, J. V. (2001). Outcome of tyrosinaemia type iii. *Journal of Inherited Metabolic Disorders, 24*(8), 824-832

Endo, F., Katoh, H., Yamamoto, S., & Matsuda, I. (1991). A murine model for type iii tyrosinemia: Lack of immunologically detectable 4-hydroxyphenylpyruvic acid dioxygenase enzyme protein in a novel mouse strain with hypertyrosinemia. *American Journal of Human Genetics, 48*(4), 704

Fang, Z., & Cui, X. (2011). Design and validation issues in rna-seq experiments. *Briefings in Bioinformatics, 12*(3), 280-287. doi:10.1093/bib/bbr004

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., . . . Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of oct-3/4 during early embryogenesis. *Nat Cell Biol, 8*(2), 188-194. doi:http://www.nature.com/ncb/journal/v8/n2/suppinfo/ncb1353_S1.html

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., & Schübeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet, 9*(12), e1003994. doi:10.1371/journal.pgen.1003994

Fellman, J. H., Fujita, T. S., & Roth, E. S. (1972). Assay, properties and tissue distribution of p-hydroxyphenylpyruvate hydroxylase. *Biochimica et Biophysica Acta, 284*(1), 90-100

Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying chip-seq enrichment using macs. *Nature protocols, 7*(9), 10.1038/nprot.2012.1101. doi:10.1038/nprot.2012.101

Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K., & Zhang, Y. (2002). Methylation of h3-lysine 79 is mediated by a new family of hmtases without a set domain. *Current Biology, 12*(12), 1052-1058

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. *Nature, 391*(6669), 806-811. doi:http://www.nature.com/nature/journal/v391/n6669/suppinfo/391806a0_S1.html

Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., . . . Allis, C. D. (2005). Regulation of hp1–chromatin binding by histone h3 methylation and phosphorylation. *Nature, 438*(7071), 1116-1122

Fowden, A. L., Forhead, A. J., Coan, P. M., & Burton, G. J. (2008). The placenta and intrauterine programming. *J Neuroendocrinol, 20*(4), 439-450. doi:10.1111/j.1365-2826.2008.01663.x

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., . . . Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences, 89*(5), 1827-1831

Frye, R. A. (1999). Characterization of five human cdnas with homology to the yeast sir2 gene: Sir2-like proteins (sirtuins) metabolize nad and may have protein adp-ribosyltransferase activity. *Biochemical and Biophysical Research Communications, 260*(1), 273-279. doi:10.1006/bbrc.1999.0897

Gao, L., Cueto, M. A., Asselbergs, F., & Atadja, P. (2002). Cloning and functional characterization of hdac11, a novel member of the human histone deacetylase family. *Journal of Biological Chemistry, 277*(28), 25748-25755

Gardin, J. M., Siri, F. M., Kitsis, R. N., Edwards, J. G., & Leinwand, L. A. (1995). Echocardiographic assessment of left ventricular mass and systolic function in mice. *Circulation Research, 76*(5), 907-914. doi:10.1161/01.res.76.5.907

Garrick, D., Fiering, S., Martin, D. I. K., & Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. *Nature Genetics, 18*(1), 56-59

Garrick, D., Sutherland, H., Robertson, G., & Whitelaw, E. (1996). Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Research, 24*(24), 4902-4909. doi:10.1093/nar/24.24.4902

Gaudet, F., Rideout, W. M., 3rd, Meissner, A., Dausman, J., Leonhardt, H., & Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of iap silencing. *Molecular and Cellular Biology, 24*(4), 1640-1648

Gautier, T., Abbott, D. W., Molla, A., Verdel, A., Ausio, J., & Dimitrov, S. (2004). Histone variant h2abbd confers lower stability to the nucleosome. *EMBO reports, 5*(7), 715-720

Gentile, J. K., Tan, W.-H., Horowitz, L. T., Bacino, C. A., Skinner, S. A., Barbieri-Welge, R., . . . Lee, H.-S. (2010). A neurodevelopmental survey of angelman syndrome with genotype-phenotype correlations. *Journal of developmental and behavioral pediatrics: JDBP, 31*(7), 592

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., . . . Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research, 15*(10), 1451-1455. doi:10.1101/gr.4086505

Ginsburg, M., Snow, M. H., & McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development, 110*(2), 521-528

Goecks, J., Nekrutenko, A., Taylor, J., & Team, T. G. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology, 11*(8), R86

Gong, S.-P., Zhao, Y.-T., & Yu, Y.-H. (2011). Vascular network modeling reveals significant differences in vascular morphology in growth-restricted placentas. *Reviews in Obstetrics and Gynecology, 4*(3-4), 103-108

Grabher, C., von Boehmer, H., & Look, A. T. (2006). Notch 1 activation in the molecular pathogenesis of t-cell acute lymphoblastic leukaemia. *Nature Reviews Cancer, 6*(5), 347-359. doi:10.1038/nrc1880

Grego-Bessa, J., Luna-Zurita, L., del Monte, G., Bolós, V., Melgar, P., Arandilla, A., . . . de la Pompa, J. L. (2007). Notch signaling is essential for ventricular chamber development. *Developmental Cell, 12*(3), 415-429. doi:10.1016/j.devcel.2006.12.011

Gregoretti, I., Lee, Y.-M., & Goodson, H. V. (2004). Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis. *Journal of Molecular Biology, 338*(1), 17-31

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., . . . Xu, G.-L. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell, 15*(4), 447-458. doi:http://dx.doi.org/10.1016/j.stem.2014.08.003

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., . . . Surani, M. A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development, 117*(1–2), 15-23. doi:10.1016/S0925-4773(02)00181-8

Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., & Surani, M. A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science, 329*(5987), 78-82. doi:10.1126/science.1187945

Hake, S. B., & Allis, C. D. (2006). Histone h3 variants and their potential role in indexing mammalian genomes: The "h3 barcode hypothesis". *Proceedings of the National Academy of Sciences, 103*(17), 6428-6435

Hang, C. T., Yang, J., Han, P., Cheng, H.-L., Shang, C., Ashley, E., . . . Chang, C.-P. (2010). Chromatin regulation by brg1 underlies heart muscle development and disease. *Nature, 466*(7302), 62-67. doi:10.1038/nature09130

Hannibal, M. C., Buckingham, K. J., Ng, S. B., Ming, J. E., Beck, A. E., McMillin, M. J., . . . Mefford, H. C. (2011). Spectrum of mll2 (alr) mutations in 110 cases of kabuki syndrome. *American Journal of Medical Genetics Part A, 155*(7), 1511-1516

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). Bsmooth: From whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol, 13*(10), R83. doi:10.1186/gb-2012-13-10-r83

Hansen, K. D., Wu, Z., Irizarry, R. A., & Leek, J. T. (2011). Sequencing technology does not eliminate biological variability. *Nature biotechnology, 29*(7), 572-573. doi:10.1038/nbt.1910

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., & Gartler, S. M. (1999). The dnmt3b DNA methyltransferase gene is mutated in the icf immunodeficiency syndrome. *Proc Natl Acad Sci U S A, 96*(25), 14412-14417

Harten, S. K., Oey, H., Bourke, L. M., Bharti, V., Isbel, L., Daxinger, L., . . . Whitelaw, E. (2015). The recently identified modifier of murine metastable epialleles, rearranged l-myc fusion, is involved in maintaining epigenetic marks at cpg island shores and enhancers. *BMC Biology, 13*(1), 21. doi:10.1186/s12915-015-0128-2

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., & Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. *Nature Genetics, 34*(2), 187-192

Hatch, C. L., & Bonner, W. M. (1988). Sequence of cdnas for mammalian h2a. Z, an evolutionarily diverged but highly conserved basal histone h2a isoprotein species. *Nucleic Acids Research, 16*(3), 1113-1124

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., . . . Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian DNA. *Science, 333*(6047), 1303-1307. doi:10.1126/science.1210944

Heard, E., & Martienssen, Robert A. (2014). Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell, 157*(1), 95-109. doi:10.1016/j.cell.2014.02.045

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., . . . Ren, B. (2009). Histone modifications at human enhancers reflect global cell type-specific gene expression. *Nature, 459*(7243), 108-112. doi:10.1038/nature07829

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors

prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell, 38*(4), 576-589. doi:10.1016/j.molcel.2010.05.004

Hemberger, M., Dean, W., & Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: Digging waddington's canal. *Nature Review Molecular Cell Biology, 10*(8), 526-537. doi:10.1038/nrm2727

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013a). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics, 45*(10), 1198-1206. doi:10.1038/ng.2746

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013b). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet, 45*(10), 1198-1206. doi:10.1038/ng.2746

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., Van Den Elst, H., . . . Moens, C. B. (2007). A role for piwi and pirnas in germ cell maintenance and transposon silencing in zebrafish. *Cell, 129*(1), 69-82

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., & Eden, A. (2007). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene, 27*(3), 404-408. doi:10.1038/sj.onc.1210631

Howlett, S. K., & Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Development, 113*(1), 119-127

Hsu, S. Y., Lai, R. J., Finegold, M., & Hsueh, A. J. (1996). Targeted overexpression of bcl-2 in ovaries of transgenic mice leads to decreased follicle apoptosis, enhanced folliculogenesis, and increased germ cell tumorigenesis. *Endocrinology, 137*(11), 4837-4843. doi:10.1210/endo.137.11.8895354

Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D., & Li, L. (2013). Pavis: A tool for peak annotation and visualization. *Bioinformatics, 29*. doi:10.1093/bioinformatics/btt520

Hübner, M. R., Eckersley-Maslin, M. A., & Spector, D. L. (2013). Chromatin organization and transcriptional regulation. *Current opinion in genetics & development, 23*(2), 89-95

Hyatsu, H., Wataya, Y., Kai, K., & Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine and their derivatives. *Biochemistry, 9*(14), 2858-2864

Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., . . . Shiota, K. (2007). Genome-wide and locus-specific DNA hypomethylation in g9a deficient mouse embryonic stem cells. *Genes to Cells, 12*(1), 1-11. doi:10.1111/j.1365-2443.2006.01029.x

Inoue, A., Shen, L., Dai, Q., He, C., & Zhang, Y. (2011). Generation and replication-dependent dilution of 5fc and 5cac during mouse preimplantation development. *Cell Res, 21*(12), 1670-1676

Inoue, A., & Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science, 334*(6053), 194-194. doi:10.1126/science.1212483

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., . . . Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science, 333*(6047), 1300-1303

Jackson, J. P., Lindroth, A. M., Cao, X., & Jacobsen, S. E. (2002). Control of cpnpg DNA methylation by the kryptonite histone h3 methyltransferase. *Nature, 416*(6880), 556-560

Jin, B., Tao, Q., Peng, J., Soo, H. M., Wu, W., Ying, J., . . . Qiu, J. (2008). DNA methyltransferase 3b (dnmt3b) mutations in icf syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Human Molecular Genetics, 17*(5), 690-709

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., & Felsenfeld, G. (2009). H3. 3/h2a. Z double variant–containing nucleosomes mark'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Genetics, 41*(8), 941-945

Johnson, M. S., Thomson, S. C., & Speakman, J. R. (2001). Limits to sustained energy intake i. Lactation in the laboratory mouse *The Journal of Experimental Biology, 204*(11), 1925-1935

Jones, P. A., Archer, T. K., Baylin, S. B., Beck, S., Berger, S., Bernstein, B. E., . . . Doerge, R. W. (2008). Moving ahead with an international human epigenome project. *Nature, 454*(7205), 711-715

Kanber, D., Giltay, J., Wieczorek, D., Zogel, C., Hochstenbach, R., Caliebe, A., . . . Buiting, K. (2008). A paternal deletion of mkrn3, magel2 and ndn does not

result in prader-willi syndrome. *Eur J Hum Genet, 17*(5), 582-590. doi:10.1038/ejhg.2008.232

Kanfi, Y., Naiman, S., Amir, G., Peshti, V., Zinman, G., Nahum, L., . . . Cohen, H. Y. (2012). The sirtuin sirt6 regulates lifespan in male mice. *Nature, 483*(7388), 218-221. doi:10.1038/nature10815

Karolchik, D., Barber, G. P., Casper, J., Clawson, H., Cline, M. S., Diekhans, M., . . . Kent, W. J. (2014). The ucsc genome browser database: 2014 update. *Nucleic Acids Res, 42*(Database issue), D764-770. doi:10.1093/nar/gkt1168

Katz, D. J., Edwards, T. M., Reinke, V., & Kelly, W. G. (2009). A c. Elegans lsd1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell, 137*(2), 308-320. doi:http://dx.doi.org/10.1016/j.cell.2009.02.015

Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N., & Rastan, S. (1993). Expression of xist during mouse development suggests a role in the initiation of x chromosome inactivation. *Cell, 72*(2), 171-182. doi:10.1016/0092-8674(93)90658-D

Kearns, M., Morris, C., & Whitelaw, E. (2001). Spontaneous germline amplification and translocation of a transgene array. *Mutation Research/DNA Repair, 486*(2), 125-136. doi:10.1016/S0921-8777(01)00084-2

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., . . . David. (2002). The human genome browser at ucsc. *Genome Research, 12*(6), 996-1006. doi:10.1101/gr.229102

Kishino, T., Lalande, M., & Wagstaff, J. (1997). Ube3a/e6-ap mutations cause angelman syndrome. *Nature Genetics, 15*(1), 70-73

Kitagawa, H., Fujiki, R., Yoshimura, K., Oya, H., & Kato, S. (2011). Williams syndrome is an epigenome-regulator disease. *Endocrine Journal, 58*(2), 77-85. doi:10.1507/endocrj.K10E-393

Kondo, T., Bobek, M. P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., . . . Hanash, S. M. (2000). Whole-genome methylation scan in icf syndrome: Hypomethylation of non-satellite DNA repeats d4z4 and nbl2. *Human Molecular Genetics, 9*(4), 597-604. doi:10.1093/hmg/9.4.597

Kornberg, R. D., & Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science, 184*(4139), 865-868

Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A., & Rossant, J. (1989). Inducible expression of an hsp68-lacz hybrid gene in transgenic mice. *Development, 105*(4), 707-714

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell, 128*(4), 693-705

Kraggerud, S. M., Hoei-Hansen, C. E., Alagaratnam, S., Skotheim, R. I., Abeler, V. M., Rajpert-De Meyts, E., & Lothe, R. A. (2013). Molecular characteristics of malignant ovarian germ cell tumors and comparison with testicular counterparts: Implications for pathogenesis. *Endocrine Reviews, 34*(3), 339-376. doi:10.1210/er.2012-1045

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics, 27*(11), 1571-1572. doi:10.1093/bioinformatics/btr167

Kumar, P., & Sait, S. F. (2011). Luteinizing hormone and its dilemma in ovulation induction. *Journal of Human Reproductive Sciences, 4*(1), 2-7. doi:10.4103/0974-1208.82351

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., . . . Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by piwi family members mili and miwi2 in murine fetal testes. *Genes & Development, 22*(7), 908-917. doi:10.1101/gad.1640708

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., . . . Ohlsson, R. (2006). Ctcf binding at the h19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to igf2. *Proceedings of the National Academy of Sciences, 103*(28), 10684-10689

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., . . . Snyder, M. (2012). Chip-seq guidelines and practices of the encode and modencode consortia. *Genome Research, 22*(9), 1813-1831. doi:10.1101/gr.136184.111

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Latham, K. E., Doherty, A. S., Scott, C. D., & Schultz, R. M. (1994). Igf2r and igf2 gene expression in androgenetic, gynogenetic, and parthenogenetic

preimplantation mouse embryos: Absence of regulation by genomic imprinting. *Genes & Development, 8*(3), 290-299. doi:10.1101/gad.8.3.290

Lau, M., Stewart, C., Liu, Z., Bhatt, H., Rotwein, P., & Stewart, C. L. (1994). Loss of the imprinted igf2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes & Development, 8*(24), 2953-2963

Ledbetter, D. H., Riccardi, V. M., Airhart, S. D., Strobel, R. J., Keenan, B. S., & Crawford, J. D. (1981). Deletions of chromosome 15 as a cause of the prader–willi syndrome. *New England Journal of Medicine, 304*(6), 325-329. doi:10.1056/NEJM198102053040604

Lederer, D., Grisart, B., Digilio, M. C., Benoit, V., Crespin, M., Ghariani, S. C., . . . Verellen-Dumoulin, C. (2012). Deletion of kdm6a, a histone demethylase interacting with mll2, in three patients with kabuki syndrome. *The American Journal of Human Genetics, 90*(1), 119-124

Lee, J., Ward, W., Knapp, G., Ren, H., Vallanat, B., Abbott, B., . . . Corton, J. (2012). Transcriptional ontogeny of the developing liver. *BMC Genomics, 13*(1), 33. doi:10.1186/1471-2164-13-33

Lee, Y., Song, A. J., Baker, R., Micales, B., Conway, S. J., & Lyons, G. E. (2000). Jumonji, a nuclear protein that is necessary for normal heart development. *Circulation Research, 86*(9), 932-938

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R., & Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development, 122*(10), 3195-3205

Lewis, A., & Reik, W. (2006). How imprinting centres work. *Cytogenetic and Genome Research, 113*(1-4), 81-89

Lewis, R. P., & Sandler, H. (1971). Relationship between changes in left ventricular dimensions and the ejection fraction in man. *Circulation, 44*(4), 548-557. doi:10.1161/01.cir.44.4.548

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell, 69*(6), 915-926

Li, G., Ruan, X., Auerbach, Raymond K., Sandhu, Kuljeet S., Zheng, M., Wang, P., . . . Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell, 148*(1–2), 84-98. doi:10.1016/j.cell.2011.12.014

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Lin, H., Gupta, V., VerMilyea, M. D., Falciani, F., Lee, J. T., O'Neill, L. P., & Turner, B. M. (2007). Dosage compensation in the mouse balances up-regulation and silencing of x-linked genes. *PLoS Biol, 5*(12), e326. doi:10.1371/journal.pbio.0050326

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., . . . Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature, 462*(7271), 315-322. doi:10.1038/nature08514

Liu, H., Hu, Q., Kaufman, A., D'Ercole, A. J., & Ye, P. (2008). Developmental expression of histone deacetylase 11 in the murine brain. *Journal of Neuroscience Research, 86*(3), 537-543

Liu, Y., Zhou, J., & White, K. P. (2014). Rna-seq differential expression studies: More sequence or more replication? *Bioinformatics, 30*(3), 301-304. doi:10.1093/bioinformatics/btt688

Livak, K., & Schmittgen, T. (2001). Analysis of relative gene expression data using real time quantitative pcr and the 2-ct method. Methods25: 402–408.

Lock, L. F., Takagi, N., & Martin, G. R. (1987). Methylation of the hprt gene on the inactive x occurs after chromosome inactivation. *Cell, 48*(1), 39-46. doi:10.1016/0092-8674(87)90353-9

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature, 389*(6648), 251-260

Luxan, G., Casanova, J. C., Martinez-Poveda, B., Prados, B., D'Amato, G., MacGrogan, D., . . . de la Pompa, J. L. (2013). Mutations in the notch pathway regulator mib1 cause left ventricular noncompaction cardiomyopathy. *Nature Medicine, 19*(2), 193-201. doi:10.1038/nm.3046

Lyon, M. F. (1961). Gene action in the x-chromosome of the mouse (mus musculus l.).

Mahler, J. F., Stokes, W., Mann, P. C., Takaoka, M., & Maronpot, R. R. (1996). Spontaneous lesions in aging fvb/n mice. *Toxicologic Pathology, 24*(6), 710-716. doi:10.1177/019262339602400606

Makela, T. P., Hellsten, E., Vesa, J., Hirvonen, H., Palotie, A., Peltonen, L., & Alitalo, K. (1995). The rearranged l-myc fusion gene (rlf) encodes a zn-15 related zinc finger protein. *Oncogene, 11*(12), 2699-2704

Makela, T. P., Kere, J., Winqvist, R., & Alitalo, K. (1991a). Intrachromosomal rearrangements fusing l-myc and rlf in small-cell lung cancer. *Molecular and Cellular Biology, 11*(8), 4015-4021

Makela, T. P., Saksela, K., Evan, G., & Alitalo, K. (1991b). A fusion protein formed by l-myc and a novel gene in sclc. *The EMBO Journal, 10*(6), 1331-1335

Makela, T. P., Shiraishi, M., Borrello, M. G., Sekiya, T., & Alitalo, K. (1992). Rearrangement and co-amplification of l-myc and rlf in primary lung cancer. *Oncogene, 7*(3), 405-409

Malcolm, S., Clayton-Smith, J., Nichols, M., Pembrey, M. E., Armour, J. A. L., Jeffreys, A. J., . . . Webb, T. (1991). Uniparental paternal disomy in angelman's syndrome. *The Lancet, 337*(8743), 694-697. doi:10.1016/0140-6736(91)90278-W

Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature Structural Biology, 10*(11), 882-891. doi:10.1038/nsb996

Malik, S., Cleves, M. A., Zhao, W., Correa, A., & Hobbs, C. A. (2007). Association between congenital heart defects and small for gestational age. *Pediatrics, 119*(4), e976-e982

Martin, D. I., & Whitelaw, E. (1996). The vagaries of variegating transgenes. *Bioessays, 18*(11), 919-923

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Embryogenesis: Demethylation of the zygotic paternal genome. *Nature, 403*(6769), 501-502

McGrath, J., & Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell, 37*(1), 179-183

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., . . . Guttman, M. (2015). The xist lncrna interacts directly with sharp to silence transcription through hdac3. *Nature, 521*(7551), 232-236. doi:10.1038/nature14443

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

McKinsey, T. A., Zhang, C.-L., Lu, J., & Olson, E. N. (2000). Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation. *Nature, 408*(6808), 106-111

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., . . . Bejerano, G. (2010). Great improves functional interpretation of cis-regulatory regions. *Nat Biotechnol, 28*(5), 495-501. doi:10.1038/nbt.1630

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research, 33*(18), 5868-5877. doi:10.1093/nar/gki901

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human argonaute2 mediates rna cleavage targeted by mirnas and sirnas. *Molecular cell, 15*(2), 185-197. doi:10.1016/j.molcel.2004.07.007

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding rnas: Insights into functions. *Nature Reviews Genetics, 10*(3), 155-159. doi:10.1038/nrg2521

Mito, Y., Henikoff, J. G., & Henikoff, S. (2005). Genome-scale profiling of histone h3. 3 replacement patterns. *Nature Genetics, 37*(10), 1090-1097

Moazed, D. (2009). Small rnas in transcriptional gene silencing and genome defence. *Nature, 457*(7228), 413-420. doi:10.1038/nature07756

Mohrmann, L., & Verrijzer, C. P. (2005). Composition and functional specificity of swi2/snf2 class chromatin remodeling complexes. *Biochimica et Biophysica Acta, 1681*(2-3), 59-73. doi:10.1016/j.bbaexp.2004.10.005

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., & Wutz, A. (2015). Identification of spen as a crucial factor for xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep, 12*(4), 554-561. doi:10.1016/j.celrep.2015.06.067

Montgomery, R. L., Davis, C. A., Potthoff, M. J., Haberland, M., Fielitz, J., Qi, X., . . . Olson, E. N. (2007). Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes & Development, 21*(14), 1790-1802

Morgan, D., & Whitelaw, E. (2008). The case for transgenerational epigenetic inheritance in humans. *Mammalian Genome, 19*(6), 394-397. doi:10.1007/s00335-008-9124-y

Morgan, H., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics, 23*(3), 314-318. doi:10.1038/15490

Moskalev, A. A., Aliper, A. M., Smit-McBride, Z., Buzdin, A., & Zhavoronkov, A. (2014). Genetics and epigenetics of aging and longevity. *Cell Cycle, 13*(7), 1063-1077. doi:10.4161/cc.28433

Mostoslavsky, R., Chua, K. F., Lombard, D. B., Pang, W. W., Fischer, M. R., Gellon, L., . . . Alt, F. W. (2006). Genomic instability and aging-like phenotype in the absence of mammalian sirt6. *Cell, 124*(2), 315-329. doi:10.1016/j.cell.2005.11.044

Muthurajan, U. M., McBryant, S. J., Lu, X., Hansen, J. C., & Luger, K. (2011). The linker region of macroh2a promotes self-association of nucleosomal arrays. *Journal of Biological Chemistry, 286*(27), 23852-23864

Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., . . . Kohli, R. M. (2012). Aid/apobec deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol, 8*(9), 751-758. doi:10.1038/nchembio.1042

Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., . . . Mefford, H. C. (2010). Exome sequencing identifies mll2 mutations as a cause of kabuki syndrome. *Nature Genetics, 42*(9), 790-793

Nguyen, D. K., & Disteche, C. M. (2006). Dosage compensation of the active x chromosome in mammals. *Nature Genetics, 38*(1), 47-53. doi:10.1038/ng1705

Nicholson, T. B., Singh, A. K., Su, H., Hevi, S., Wang, J., Bajko, J., . . . Chen, T. (2013). A hypomorphic lsd1 allele results in heart development defects in mice. *PLoS One, 8*(4), e60913. doi:10.1371/journal.pone.0060913

Niederwieser, A., Matasovic, A., Tippett, P., & Danks, D. M. (1977). A new sulfur amino acid, named hawkinsin, identified in a baby with transient tyrosinemia and her mother. *Clinical Chimica Acta, 76*(3), 345-356

Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T., & Kajii, T. (1981). Kabuki make-up syndrome: A syndrome of mentalretardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *The Journal of pediatrics, 99*(4), 565-569

Norouzitallab, P., Baruah, K., Vandegehuchte, M., Van Stappen, G., Catania, F., Vanden Bussche, J., . . . Bossier, P. (2014). Environmental heat stress induces epigenetic transgenerational inheritance of robustness in parthenogenetic artemia model. *The FASEB Journal, 28*(8), 3552-3563. doi:10.1096/fj.14-252049

North, J. A., Javaid, S., Ferdinand, M. B., Chatterjee, N., Picking, J. W., Shoffner, M., . . . Poirier, M. G. (2011). Phosphorylation of histone h3(t118) alters nucleosome dynamics and remodeling. *Nucleic Acids Research, 39*(15), 6465-6474. doi:10.1093/nar/gkr304

Nozawa, R.-S., Nagao, K., Masuda, H.-T., Iwasaki, O., Hirota, T., Nozaki, N., . . . Obuse, C. (2010). Human pogz modulates dissociation of hp1α from mitotic chromosome arms through aurora b activation. *Nature cell biology, 12*(7), 719-727

O'Meara, C. C., Wamstad, J. A., Gladstone, R. A., Fomovsky, G. M., Butty, V. L., Shrikumar, A., . . . Lee, R. T. (2015). Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. *Circulation Research, 116*(5), 804-815. doi:10.1161/circresaha.116.304269

Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., . . . Baylin, S. B. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genetics, 39*(2), 237-242. doi:10.1038/ng1972

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell, 99*(3), 247-257

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics, 19*(3), 219-220. doi:10.1038/890

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., . . . Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology, 10*(8), 475-478

Park, J. H., Park, E. J., Lee, H. S., Kim, S. J., Hur, S. K., Imbalzano, A. N., & Kwon, J. (2006). Mammalian swi/snf complexes facilitate DNA double-strand break repair by promoting gamma-h2ax induction. *The EMBO Journal, 25*(17), 3986-3997. doi:10.1038/sj.emboj.7601291

Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M., & Bonner, W. M. (2000). A critical role for histone h2ax in recruitment of repair factors to nuclear foci after DNA damage. *Current Biology, 10*(15), 886-895

Peltoketo, H., Strauss, L., Karjalainen, R., Zhang, M., Stamp, G. W., Segaloff, D. L., . . . Huhtaniemi, I. T. (2010). Female mice expressing constitutively active mutants of fsh receptor present with a phenotype of premature follicle depletion and estrogen excess. *Endocrinology, 151*(4), 1872-1883. doi:10.1210/en.2009-0966

en.2009-0966 [pii]

Pembrey, M. E., Bygren, L. O., Kaati, G., Edvinsson, S., Northstone, K., Sjostrom, M., & Golding, J. (2005). Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet, 14*(2), 159-166. doi:10.1038/sj.ejhg.5201538

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., . . . Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature, 444*(7118), 499-502. doi:10.1038/nature05295

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for xist in x chromosome inactivation. *Nature, 379*(6561), 131-137. doi:10.1038/379131a0

Percec, I., Plenge, R. M., Nadeau, J. H., Bartolomei, M. S., & Willard, H. F. (2002). Autosomal dominant mutations affecting x inactivation choice in the mouse. *Science, 296*(5570), 1136-1139. doi:10.1126/science.1070087

Peters, A. H., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A., Perez-Burgos, L., . . . Shinkai, Y. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular cell, 12*(6), 1577-1589

Polin, R. A., Fox, W. W., & Abman, S. H. (2004). *Fetal and neonatal physiology* (Vol. 3rd). Philadelphia: W.B. Saunders Co.

Preis, J. I., Downes, M., Oates, N. A., Rasko, J. E., & Whitelaw, E. (2003). Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. *Current Biology, 13*(11), 955-959

Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I., & Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends in Genetics, 18*(7), 348-351

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (2000). Non-cpg methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences, 97*(10), 5237-5242

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., . . . Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for rna-seq data. *Genome Biology, 14*(9), R95

Rasko, J. E., Battini, J.-L., Gottschalk, R. J., Mazo, I., & Miller, A. D. (1999). The rd114/simian type d retrovirus receptor is a neutral amino acid transporter. *Proceedings of the National Academy of Sciences, 96*(5), 2129-2134

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., . . . Allis, C. D. (2000). Regulation of chromatin structure by site-specific histone h3 methyltransferases. *Nature, 406*(6796), 593-599

Remus, R., Kämmer, C., Heller, H., Schmitz, B., Schell, G., & Doerfler, W. (1999). Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. *Journal of Virology, 73*(2), 1010-1022

Richards, J. S., & Pangas, S. A. (2010). The ovary: Basic biology and clinical implications. *The Journal of Clinical Investigation, 120*(4), 963-972. doi:10.1172/JCI41350

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., . . . Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human hox loci by noncoding rnas. *Cell, 129*(7), 1311-1323. doi:http://dx.doi.org/10.1016/j.cell.2007.05.022

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). Edger: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140

Rogakou, E. P., Boon, C., Redon, C., & Bonner, W. M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *The Journal of Cell Biology, 146*(5), 905-916

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., . . . Kent, W. J. (2013a). Encode data in the ucsc genome browser:

Year 5 update. *Nucleic Acids Res, 41*(Database issue), D56-63. doi:10.1093/nar/gks1172

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., . . . Kent, W. J. (2013b). Encode data in the ucsc genome browser: Year 5 update. *Nucleic Acids Research, 41*(D1), D56-D63. doi:10.1093/nar/gks1172

Rosenfeld, J. A., Wang, Z., Schones, D. E., Zhao, K., DeSalle, R., & Zhang, M. Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics, 10*(1), 143

Rossant, J., & Cross, J. C. (2001). Placental development: Lessons from mouse mutants. *Nature Reviews Genetics, 2*(7), 538-548. doi:10.1038/35080570

Rudin, C. M., Durinck, S., Stawiski, E. W., Poirier, J. T., Modrusan, Z., Shames, D. S., . . . Seshagiri, S. (2012). Comprehensive genomic analysis identifies sox2 as a frequently amplified gene in small-cell lung cancer. *Nature Genetics, 44*(10), 1111-1116. doi:10.1038/ng.2405

Sambrook, J., & Russell, D. W. (2001). Molecular cloning. A laboratory manual. Third. *Cold pring Harbor Laboratory Press, New York*

Santen, G. W. E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., . . . Kriek, M. (2012). Mutations in swi/snf chromatin remodeling complex gene arid1b cause coffin-siris syndrome. *Nature Genetics, 44*(4), 379-380. doi:10.1038/ng.2217

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of cpg dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America, 103*(5), 1412-1417. doi:10.1073/pnas.0510310103

Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeier, C., Truss, M., & Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie International Edition, 51*(26), 6516-6520. doi:10.1002/anie.201202583

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., . . . Cedar, H. (2007). Polycomb-mediated methylation on lys27 of histone h3 pre-marks genes for de novo methylation in cancer. *Nature Genetics, 39*(2), 232-236. doi:10.1038/ng1950

Schotta, G., Ebert, A., Dorn, R., & Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in drosophila. *Seminars in Cell & Developmental Biology, 14*(1), 67-75. doi:10.1016/S1084-9521(02)00138-6

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., . . . Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature, 488*(7409), 116-120. doi:10.1038/nature11243

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., . . . Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog lsd1. *Cell, 119*(7), 941-953

Shin, J. Y., Fitzpatrick, G. V., & Higgins, M. J. (2008). Two distinct mechanisms of silencing by the kvdmr1 imprinting control region. *The EMBO Journal, 27*(1), 168-178

Sibley, C. P., Turner, M. A., Cetin, I., Ayuk, P., Boyd, C. A. R., D'Souza, S. W., . . . Powell, T. (2005). Placental phenotypes of intrauterine growth. *Pediatric Research, 58*(5), 827-832. doi:10.1203/01.PDR.0000181381.82856.23

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., . . . Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res, 15*(8), 1034-1050. doi:10.1101/gr.3715005

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: A germ cell perspective. *Trends in Genetics, 28*(1), 33-42. doi:10.1016/j.tig.2011.09.004

Soboleva, T. A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G. A., & Tremethick, D. J. (2012). A unique h2a histone variant occupies the transcriptional start site of active genes. *Nature structural & molecular biology, 19*(1), 25-30

Sousa, S. B., Abdul-Rahman, O. A., Bottani, A., Cormier-Daire, V., Fryer, A., Gillessen-Kaesbach, G., . . . Hennekam, R. C. M. (2009). Nicolaides–baraitser syndrome: Delineation of the phenotype. *American Journal of Medical Genetics Part A, 149A*(8), 1628-1640. doi:10.1002/ajmg.a.32956

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., . . . Gaidatzis, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*

Stankunas, K., Hang, C. T., Tsun, Z. Y., Chen, H., Lee, N. V., Wu, J. I., . . . Chang, C. P. (2008). Endocardial brg1 represses adamts1 to maintain the microenvironment for myocardial morphogenesis. *Dev Cell, 14*(2), 298-311. doi:10.1016/j.devcel.2007.11.018

Stefansson, O. A., Moran, S., Gomez, A., Sayols, S., Arribas-Jorba, C., Sandoval, J., . . . Esteller, M. (2015). A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Molecular Oncology, 9*(3), 555-568. doi:10.1016/j.molonc.2014.10.012

Stopka, T., & Skoultchi, A. I. (2003). The iswi atpase snf2h is required for early mouse development. *Proceedings of the National Academy of Sciences of the United States of America, 100*(24), 14097-14102. doi:10.1073/pnas.2336105100

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature, 403*(6765), 41-45

Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Langst, G., & Grummt, I. (2001). Norc[mdash]a novel member of mammalian iswi-containing chromatin remodeling machines. *The EMBO Journal, 20*(17), 4892-4900. doi:10.1093/emboj/20.17.4892

Surani, M., Barton, S., & Norris, M. (1986). Nuclear transplantation in the mouse: Heritable differences between parental genomes after activation of the embryonic genome. *Cell, 45*(1), 127-136

Suto, R. K., Clarkson, M. J., Tremethick, D. J., & Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone h2a. Z. *Nature structural & molecular biology, 7*(12), 1121-1124

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., & Shinkai, Y. (2008). G9a/glp complexes independently mediate h3k9 and DNA methylation to silence transcription. *The EMBO Journal, 27*(20), 2681-2690. doi:10.1038/emboj.2008.192

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., . . . Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by mll partner tet1. *Science, 324*(5929), 930-935. doi:10.1126/science.1170116

Tamaru, H., & Selker, E. U. (2001). A histone h3 methyltransferase controls DNA methylation in neurospora crassa. *Nature, 414*(6861), 277-283

Tanaka, N., Dalton, N., Mao, L., Rockman, H. A., Peterson, K. L., Gottshall, K. R., . . . Ross, J. (1996). Transthoracic echocardiography in models of cardiac disease in the mouse. *Circulation, 94*(5), 1109-1117. doi:10.1161/01.cir.94.5.1109

Tanaka, Y., Nakamura, K., Matsumoto, S., Kimoto, Y., Tanoue, A., Tsujimoto, G., & Endo, F. (2006). Gene expression profiles of homogentisate-treated fah-/-hpd-/-mice using DNA microarrays. *Molecular Genetics and Metabolism, 89*(3), 203-209. doi:10.1016/j.ymgme.2005.09.022

Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator rpd3p. *Science, 272*, 408+

Thatcher, T. H., & Gorovsky, M. A. (1994). Phylogenetic analysis of the core histones h2a, h2b, h3, and h4. *Nucleic Acids Research, 22*(2), 174-179

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi:10.1038/nature11247

Thornburg, K. L., O'Tierney, P. F., & Louey, S. (2010). Review: The placenta is a programming agent for cardiovascular disease. *Placenta, 31, Supplement*(0), S54-S59. doi:10.1016/j.placenta.2010.01.002

Tjeertes, J. V., Miller, K. M., & Jackson, S. P. (2009). Screen for DNA-damage-responsive histone modifications identifies h3k9ac and h3k56ac in human cells. *The EMBO Journal, 28*(13), 1878-1889

Tolstorukov, M. Y., Goldman, J. A., Gilbert, C., Ogryzko, V., Kingston, R. E., & Park, P. J. (2012). Histone variant h2a. Bbd is associated with active transcription and mrna processing in human cells. *Molecular cell, 47*(4), 596-607

Tomoeda, K., Awata, H., Matsuura, T., Matsuda, I., Ploechl, E., Milovac, T., . . . Endo, F. (2000). Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type iii and hawkinsinuria. *Molecular Genetics and Metabolism, 71*(3), 506-510

Tost, J., & Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. *Nature protocols, 2*(9), 2265-2275. doi:10.1038/nprot.2007.314

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). Tophat: Discovering splice junctions with rna-seq. *Bioinformatics, 25*(9), 1105-1111

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., . . . Pachter, L. (2012). Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nature protocols, 7*(3), 562-578

Tsai, T.-F., Chen, K.-S., Weber, J. S., Justice, M. J., & Beaudet, A. L. (2002). Evidence for translational regulation of the imprinted snurf–snrpn locus in mice. *Human Molecular Genetics, 11*(14), 1659-1668. doi:10.1093/hmg/11.14.1659

Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., & Zhang, Y. (2006). Histone demethylation by a family of jmjc domain-containing proteins. *Nature, 439*(7078), 811-816

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S.-i., Sakaue, M., Matsuoka, C., . . . Okano, M. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases dnmt1, dnmt3a and dnmt3b. *Genes to Cells, 11*(7), 805-814. doi:10.1111/j.1365-2443.2006.00984.x

Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., . . . Matsumoto, N. (2012). Mutations affecting components of the swi/snf complex cause coffin-siris syndrome. *Nature Genetics, 44*(4), 376-378. doi:10.1038/ng.2219

Valinluck, V., & Sowers, L. C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase dnmt1. *Cancer Research, 67*(3), 946-950. doi:10.1158/0008-5472.can-06-3123

Van Laarhoven, P. M., Neitzel, L. R., Quintana, A. M., Geiger, E. A., Zackai, E. H., Clouthier, D. E., . . . Shaikh, T. H. (2015). Kabuki syndrome genes kmt2d and kdm6a: Functional analyses demonstrate critical roles in craniofacial, heart and brain development. *Human Molecular Genetics, 24*(15), 4443-4453. doi:10.1093/hmg/ddv180

Vega, R. B., Harrison, B. C., Meadows, E., Roberts, C. R., Papst, P. J., Olson, E. N., & McKinsey, T. A. (2004). Protein kinases c and d mediate agonist-dependent cardiac hypertrophy through nuclear export of histone deacetylase 5. *Molecular and Cellular Biology, 24*(19), 8374-8385. doi:10.1128/mcb.24.19.8374-8385.2004

Vicente-Steijn, R., Scherptong, R. W. C., Kruithof, B. P. T., Duim, S. N., Goumans, M. J. T. H., Wisse, L. J., . . . Jongbloed, M. R. M. (2015). Regional

differences in wt-1 and tcf21 expression during ventricular development: Implications for myocardial compaction. *PLoS One, 10*(9), e0136025. doi:10.1371/journal.pone.0136025

Vignali, M., & Workman, J. L. (1998). Location and function of linker histones. *Nature structural & molecular biology, 5*(12), 1025-1028. doi:10.1038/4133

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). Vista enhancer browser—a database of tissue-specific human enhancers. *Nucleic Acids Research, 35*(suppl 1), D88-D92. doi:10.1093/nar/gkl822

Walter, W., Sanchez-Cabo, F., & Ricote, M. (2015). Goplot: An r package for visually combining expression data with functional analysis. *Bioinformatics, 31*(17), 2912-2914. doi:10.1093/bioinformatics/btv300

Wang, Z., Zhai, W., Richardson, J. A., Olson, E. N., Meneses, J. J., Firpo, M. T., . . . Tjian, R. (2004). Polybromo protein baf180 functions in mammalian cardiac chamber maturation. *Genes & Development, 18*(24), 3106-3116. doi:10.1101/gad.1238104

Waterland, R. A., Travisano, M., & Tahiliani, K. G. (2007). Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *The FASEB Journal, 21*(12), 3380-3385. doi:10.1096/fj.07-8229com

Watson, E. D., & Cross, J. C. (2005). Development of structures and transport functions in the mouse placenta. *Physiology, 20*(3), 180-193. doi:10.1152/physiol.00001.2005

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., & Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics, 37*(8), 853-862. doi:10.1038/ng1598

Wijmenga, C., Hansen, R. S., Gimelli, G., Bjorck, E. J., Davies, E. G., Valentine, D., . . . Pearson, P. L. (2000). Genetic variation in icf syndrome: Evidence for genetic heterogeneity. *Hum Mutat, 16*(6), 509-517. doi:10.1002/1098-1004(200012)16:6<509::aid-humu8>3.0.co;2-v

Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., . . . Schultz, P. G. (2005). A strategy for probing the function of noncoding rnas finds a repressor of nfat. *Science, 309*(5740), 1570-1573. doi:10.1126/science.1115901

Wolff, G. L., Kodell, R. L., Moore, S. R., & Cooney, C. A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in avy/a mice. *Faseb j, 12*(11), 949-957

Xhemalce, B., Dawson, M. A., & Bannister, A. J. (2006). Histone modifications *Reviews in cell biology and molecular medicine*: Wiley-VCH Verlag GmbH & Co. KGaA.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., . . . Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature, 402*(6758), 187-191. doi:10.1038/46052

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Côté, J., & Wang, W. (1998). Nurd, a novel complex with both atp-dependent chromatin-remodeling and histone deacetylase activities. *Molecular cell, 2*(6), 851-861. doi:10.1016/S1097-2765(00)80299-3

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., . . . Tang, F. (2013). Single-cell rna-seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology, 20*(9), 1131-1139. doi:10.1038/nsmb.2660

Yang, J., Bücker, S., Jungblut, B., Böttger, T., Cinnamon, Y., Tchorz, J., . . . Braun, T. (2012). Inhibition of notch2 by numb/numblike controls myocardial compaction in the heart. *Cardiovascular Research, 96*(2), 276-285. doi:10.1093/cvr/cvs250

Yang, X. J., & Seto, E. (2007). Hats and hdacs: From structure, function and regulation to novel strategies for therapy and prevention. *Oncogene, 26*(37), 5310-5318. doi:10.1038/sj.onc.1210599

Yang, Y., Fear, J., Hu, J., Haecker, I., Zhou, L., Renne, R., . . . McIntyre, L. M. (2014). Leveraging biological replicates to improve analysis in chip-seq experiments. *Computational and Structural Biotechnology Journal, 9*(13), 1-10. doi:10.5936/csbj.201401002

Yin, S., Deng, W., Zheng, H., Zhang, Z., Hu, L., & Kong, X. (2009). Evidence that the nonsense-mediated mrna decay pathway participates in x chromosome dosage compensation in mammals. *Biochemical and Biophysical Research Communications, 383*(3), 378-382. doi:10.1016/j.bbrc.2009.04.021

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., & Okazaki, T. (2000). Human centromere protein a (cenp-a) can replace histone h3 in nucleosome reconstitution in vitro. *Proceedings of the National Academy of Sciences, 97*(13), 7266-7271

You, A., Tong, J. K., Grozinger, C. M., & Schreiber, S. L. (2001). Corest is an integral component of the corest- human histone deacetylase complex. *Proceedings of the National Academy of Sciences of the United States of America, 98*(4), 1454-1458

Youngson, N., Epp, T., Roberts, A., Daxinger, L., Ashe, A., Huang, E., . . . Whitelaw, E. (2013). No evidence for cumulative effects in a dnmt3b hypomorph across multiple generations. *Mammalian Genome, 24*(5-6), 206-217. doi:10.1007/s00335-013-9451-5

Youngson, N. A., Vickaryous, N., van der Horst, A., Epp, T., Harten, S., Fleming, J. S., . . . Whitelaw, E. (2011). A missense mutation in the transcription factor foxo3a causes teratomas and oocyte abnormalities in mice. *Mammalian Genome, 22*(3-4), 235-248. doi:10.1007/s00335-011-9317-7

Youngson, N. A., & Whitelaw, E. (2008). Transgenerational epigenetic effects. *Annual Review of Genomics and Human Genetics, 9*, 233-257. doi:10.1146/annurev.genom.9.081307.164445

Yuen, R. K. C., Penaherrera, M. S., von Dadelszen, P., McFadden, D. E., & Robinson, W. P. (2010). DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *Eur J Hum Genet, 18*(9), 1006-1012. doi:10.1038/ejhg.2010.63

Zeller, C., Dai, W., Steele, N. L., Siddiq, A., Walley, A. J., Wilhelm-Benartzi, C. S. M., . . . Brown, R. (2012). Candidate DNA methylation drivers of acquired cisplatin resistance in ovarian cancer identified by methylome and expression profiling. *Oncogene, 31*(42), 4567-4576. doi:10.1038/onc.2011.611

Zerbino, D., Wilder, S., Johnson, N., Juettemann, T., & Flicek, P. (2015). The ensembl regulatory build. *Genome Biology, 16*(1), 56. doi:10.1186/s13059-015-0621-5

Zhang, J., Poh, H. M., Peh, S. Q., Sia, Y. Y., Li, G., Mulawadi, F. H., . . . Ruan, Y. (2012a). Chia-pet analysis of transcriptional chromatin interactions. *Methods, 58*(3), 289-299. doi:10.1016/j.ymeth.2012.08.009

Zhang, W., Zhang, T., Wu, Y., & Jiang, J. (2012b). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in arabidopsis. *Plant Cell, 24*(7), 2719-2731. doi:10.1105/tpc.112.098061

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of chip-seq (macs). *Genome Biology, 9*. doi:10.1186/gb-2008-9-9-r137

Zheng, X., Chen, L., Li, M., Lou, Q., Xia, H., Wang, P., . . . Luo, L. (2013). Transgenerational variations in DNA methylation induced by drought stress in two rice varieties with distinguished difference to drought resistance. *PLoS One, 8*(11), e80253. doi:10.1371/journal.pone.0080253

Zhou, B., Ma, Q., Rajagopal, S., Wu, S. M., Domian, I., Rivera-Feliciano, J., . . . Pu, W. T. (2008). Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature, 454*(7200), 109-113. doi:10.1038/nature07060

Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nature Methods, 12*(3), 230-232. doi:10.1038/nmeth.3152

Zlatanova, J., & Thakar, A. (2008). H2a. Z: View from the top. *Structure, 16*(2), 166-179